

PROTEOIQ

BEYOND PROTEIN IDENTIFICATIONS

by

BIOINQUIRE

User Guide

Version 1.5

Release date 1/11/10

Table of Contents

Getting Started.....	6
Tutorials and Quick Start Guide.....	6
System Requirements	7
Installing ProteoIQ.....	8
Help files and sample data	8
Licensing	8
Contact Information	9
ProteoIQ Basics	10
Navigation Panel	11
New Project Wizard – Label Free Quantitation	12
New Project Wizard – Reporter Ion Quantitation.....	13
Modify Database Wizard.....	14
Peptide/Protein Viewer	15
Protein Set Navigation Bar	16
Protein Set Viewer	17
Protein Sequence Viewer	18
Peptide Set Viewer	20
MS/MS Spectra Viewer.....	21
Protein Set Parameters Field and Output Notes	22
Menu Bar	23
File Menu	23
Edit Menu	24
View Menu.....	25
Plots Menu.....	26
Help Menu	28
Pop-Up-Menus.....	28
Statistical Algorithms Overview	29

Data File Formats.....	30
How to Use ProteoIQ.....	31
Opening a ProteoIQ Project	32
Saving a ProteoIQ Project	33
Sharing and Moving a ProteoIQ Project	35
Creating a New Project Overview	36
What You Need to Create a New Project	37
Creating a New Project – Label Free Quantitation	39
Describe Project.....	40
Describe Proteome	41
Describe Samples.....	43
Choose Database	44
Search Results Options	44
Select Files for Parsing.....	46
Peptide Parsing Options	47
Parse Files\MS Ions Data	50
Cluster Peptides.....	51
Creating a New Project – Reporter Ion Quantitation	52
Describe Experiment	53
Describe Correction Factor.....	55
Describe Samples.....	56
Map Ions to Groups	57
Select Files for Parsing.....	58
Select Label Modification	59
Modify Database Wizard.....	60
Describe Database	61
Specify Parse Rules	62
Test Parse Rules.....	64
Specify Hyperlinks.....	65
Using the Filter Pane.....	66
Selecting Filter Panes.....	66
How Filtering Works	67

Selecting Protein Sets to Filter	68
Description of Filter Parameters	69
Comparing Protein Sets	73
Customizing Results Views	75
Moving and Sorting Columns	75
Peptide Table Options in Peptide Set Viewer.....	76
Protein Table Options in Protein Set Viewer	79
Customize Protein and Peptide Tables in Protein Sequence Viewer	83
Modifying Plot Views	84
Select Proteins In View	84
Exporting Results	85
Protein Sets to Export.....	85
Data to Export.....	86
Parameters for Creation of Spectral Images	87
Evaluating Results	89
Peptide Identification	89
Score Filters	90
Peptide Probabilities	91
Peptide False Discovery Rates	93
Validating MS/MS Spectra.....	95
Spectral Viewer Overview	95
Adjusting Peak Labeling.....	96
Mapping Neutral Losses	98
Validating Reporter Ion Quantitation.....	99
Protein Identification	100
Protein Score Filters	100
Protein Probability.....	101
Protein False Discovery Rate	102
Whole Proteome Visualization	104
Pie Chart	105
Venn Diagram - Group Summary (Venn).....	106
Traditional Venn Diagram Labeling	107

Strict Venn Diagram Labeling	107
2D Gel View	108
1D Gel View	109
Experimental Approach for GeLC Analysis	109
Preparing results for GeLC Analysis in ProteoIQ	109
How to display a 1D Gel View.....	110
Interpreting the 1D Gel View.....	110
1D Gel View by Protein.....	111
Quantitation.....	112
Quantitation Overview – Label Free.....	113
Quantitation Without Normalization – Label Free.....	115
Quantitation With Normalization – Label Free	115
Normalization by Selected Proteins	116
Sample Normalization	117
Protein Size Normalization (NSAF)	119
Modifying Label Free Quantitation Settings.....	120
Quantitation Overview – Reporter Ion Quantitation	121
The ProteoIQ Quantitation Workflow – Reporter Ion Quantitation	122
Modifying Reporter Quantitation Settings.....	123
Pseudo Spectral Counts	124
The Pseudo Spectral Count Workflow.....	125
Pseudo Spectral Counts Without Normalization.....	125
Pseudo Spectral Counts With Normalization	126
Analyzing Replicate Results	127
Replicate Filters	127
Replicate Plots	128
Visualizing Quantitative Results	129
Quantitative Expression (Scatter).....	129
Quantitative Expression (Heat Map)	131
Relative Expression Changes (Cluster Dot).....	133
Relative Expression Differences (Dot)	135
Other Topics.....	137

Protein Grouping	138
Source	139
Redundancy and Multiple Charge States	140
Discriminant Score (F value)	141
Legal Notices	142
Index	144

Getting Started

The ProteoIQ user guide includes information about the use of ProteoIQ version 1.3 for the analysis of proteomics results derived from database searching of MS/MS data.

Contents

- 🔍 [Tutorials and Quick Start Guide](#)
- 🔍 [System Requirements](#)
- 🔍 [Installing ProteoIQ](#)
- 🔍 [Licensing ProteoIQ](#)
- 🔍 [Contact Information](#)

Tutorials and Quick Start Guide

PDF versions of the tutorials, help guide, and quick start guide can be found in the ProteoIQ installation path at C:\Program files\ProteoIQ\docs.

These documents can also be accessed online at: <https://www.bioinquire.com/library.php>.

Tutorials

1. Database Wizard Guide
2. How to Create Hyperlinks
3. Guide to Grouping Database Search Results
4. Managing Protein Sets

Onboard help is also available in ProteoIQ. Select **HELP** to access the onboard help guide.

System Requirements

This release of ProteoIQ requires Java version 1.6.14 or higher. ProteoIQ will not work on earlier versions of Java. If Java version 1.6.14 is not installed, ProteoIQ will automatically install it during the installation process. ProteoIQ can be operated on Windows, Mac, or Linux operating systems.

Less than 400,000 spectra in a single ProteoIQ project	Minimum Requirements	Recommended System Configuration
Processor	1.5 GHz Pentium IV 32-bit (x86) or better	2.0 GHz 32-bit (x86) or better
System Type	32-bit	32-bit
Memory	1.5 GB or more	2 GB or more
Disk Space	200 MB for installation 2 GB to run	200 MB for installation 4 GB to run
Internet Connection**	Cable or DSL	Cable or DSL

More than 400,000 spectra in a single ProteoIQ project	Minimum Requirements	Recommended System Configuration
Processor	2.0 GHz 64 bit (x64) or better	2.0 GHz 64 bit (x64) or better
System Type	64-bit	64-bit
Memory*	2 GB or more*	4 GB or more*
Disk Space	200 MB for installation 4 GB to run	200 MB for installation 8 GB or more to run***
Internet Connection**	Cable or DSL	Cable or DSL

*The number of spectra in a single project is scalable with the amount of RAM

**The computer must be attached to the internet during installation for license validation. For limited use licenses (ProteoIQ 40 and 200) the computer must have access to the internet each time the software is used.

***Amount of disk space required for is at least twice the amount of RAM (e.g. 4GB requires at least 8GB disk space)

Installing ProteoIQ

Double click on the ProteoIQ installer and follow the instructions to install ProteoIQ. For Windows Vista users, you may need to right click on the ProteoIQ installer and select "Run as administrator".

The default installation path for ProteoIQ is:

C:\Program files\ProteoIQ\

Help files and sample data

The help files and the quick start guide can be found in:

C:\Program files\ProteoIQ\docs

To locate sample data files go to:

C:\Program files\ProteoIQ\Sample data\Comparative Mouse Cell Proteome

To open the sample data select "Open Project" from the main toolbar, browse for the above folder then select "validated_Mouse cell proteome.piq".

Licensing

From the ProteoIQ navigation panel select View License. To enter a new license type the license code into the fields available and select submit. **Note:** Your computer must have an active internet connection. For unlimited licenses this will be the only time ProteoIQ has to confirm the license so your computer can be removed from the internet once the license has been activated. If you are using the ProteoIQ 40 or 200 versions the license key must be confirmed each time ProteoIQ is opened so the computer must always have access to the internet.

If your license has expired or you are using the ProteoIQ viewer select Open Viewer Only. This will allow you to access previously processed ProteoIQ data files without the need for an active license key.

Contact Information

For technical support and general inquiries

Phone: 706-583-4000

Fax: 706-583-4037

Email: support@bioinquire.com

Online support: <https://www.bioinquire.com/library.php>

For sales or to obtain a license

Phone: 706-583-4000

Fax: 706-583-4037

Email: sales@bioinquire.com

To purchase online: <https://www.bioinquire.com/editions-and-pricing.php>

ProteoIQ Basics

ProteoIQ version 1.5 supports the analysis of database search results from Mascot, SEQUEST, and X!Tandem. The software is built using a modular design in which all windows can be opened separately and two wizards guide you through the process of importing the results into ProteoIQ. This chapter provides an overview of the interfaces and functionality in ProteoIQ.

Contents

- 🔍 [Navigation panel](#)
- 🔍 [New Project Wizard – Label Free Quantitation](#)
- 🔍 [New Project Wizard – Reporter Quantitation](#)
- 🔍 [Modify Database Wizard](#)
- 🔍 [Peptide/Protein Set Viewer](#)
- 🔍 [Statistical Algorithms Overview](#)
- 🔍 [Data File Formats](#)

Navigation Panel

Figure 1 shows the ProteoIQ Navigation panel. When ProteoIQ is launched, the navigation panel is docked to the left side of the screen. The following items can be accessed via the navigation panel.

- 🔍 [New Project Wizard](#)
- 🔍 [Open Project](#)
- 🔍 [Modify Database Wizard](#)
- 🔍 [License Manager](#)
- 🔍 [Help Guide](#)
- 🔍 **Minimize and Close**

Figure 1. ProteoIQ Navigation Panel



Note: You can open multiple windows simultaneously. For example, an existing project can be viewed while a New Project is being created.

New Project Wizard – Label Free Quantitation

Figure 2 shows the ProteoIQ New Project Wizard for Label Free Quantitation. The New Project Wizard is used to create a new ProteoIQ project. Step-by-step instructions are provided to assist the user in grouping and importing database search results. Navigating through the New Project Wizard occurs in thirteen steps.

- 🔍 Describe Project
- 🔍 Describe Proteome
- 🔍 Describe Samples
- 🔍 Choose Database
- 🔍 Search Result Options
- 🔍 Select Target Files for Parsing
- 🔍 Select Decoy Files for Parsing
- 🔍 Select Non Target Files for Parsing
- 🔍 Peptide Parsing Options
- 🔍 Parse Files
- 🔍 Parse MS Ion Data
- 🔍 Cluster Peptides
- 🔍 Peptide Protein Viewer

Figure 2. New Project Wizard – Label Free Quantitation

Wizard provides step-by-step instructions for importing database search results into ProteoIQ

Define parameters for project creation

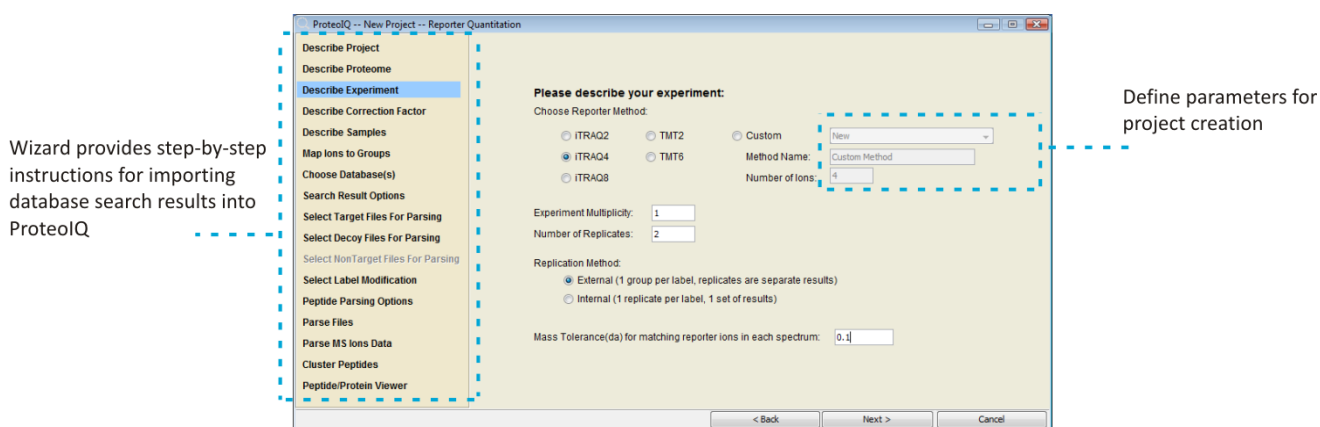
Note: For more information on grouping database search results please see the “Grouping database search results” tutorial.

New Project Wizard – Reporter Ion Quantitation

Figure 3 shows the ProteoIQ New Project Wizard for Reporter Ion Quantitation. The New Project Wizard is used to create a new ProteoIQ project. Step-by-step instructions are provided to assist the user in grouping and importing database search results. Navigating through the New Project Wizard occurs in seventeen steps.

- Q Describe Project
- Q Describe Proteome
- Q Described Experiment
- Q Describe Correction Factor
- Q Describe Samples
- Q Map Ions to Groups
- Q Choose Database
- Q Search Result Options
- Q Select Target Files for Parsing
- Q Select Decoy Files for Parsing
- Q Select Non Target Files for Parsing
- Q Select Label Modification
- Q Peptide Parsing Options
- Q Parse Files
- Q Parse MS Ion Data
- Q Cluster Peptides
- Q Peptide Protein Viewer

Figure 3. New Project Wizard – Reporter Ion Quantitation

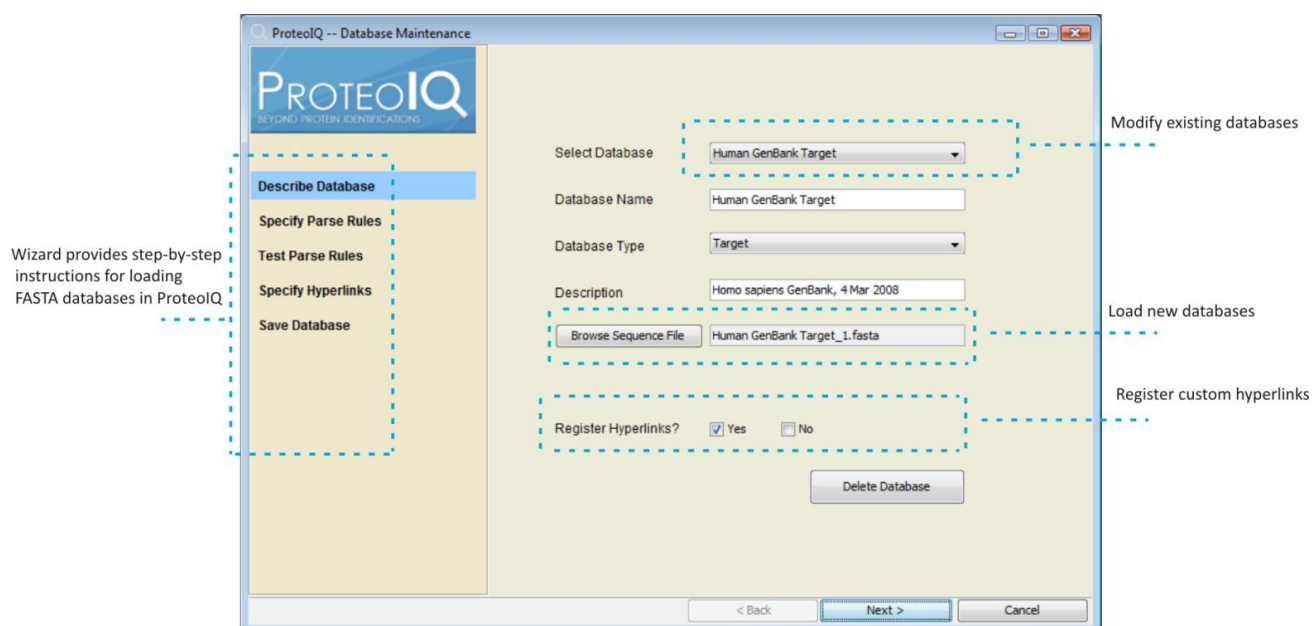


Modify Database Wizard

Figure 4 shows the ProteoIQ Modify Database Wizard. The Modify Database Wizard enables FASTA databases to be saved in ProteoIQ. Previously saved databases can be modified within the wizard and custom hyperlinks can be registered to link protein identifications to external information resources. There are five components to the Modify Database Wizard.

- 🔍 [Describe Database](#)
- 🔍 [Specify Parse Rules](#)
- 🔍 [Test Parse Rules](#)
- 🔍 [Specify Hyperlinks](#)
- 🔍 **Save Database**

Figure 4. Modify Database Wizard



Note: For more information on loading FASTA databases please see the “Database Wizard” tutorial.

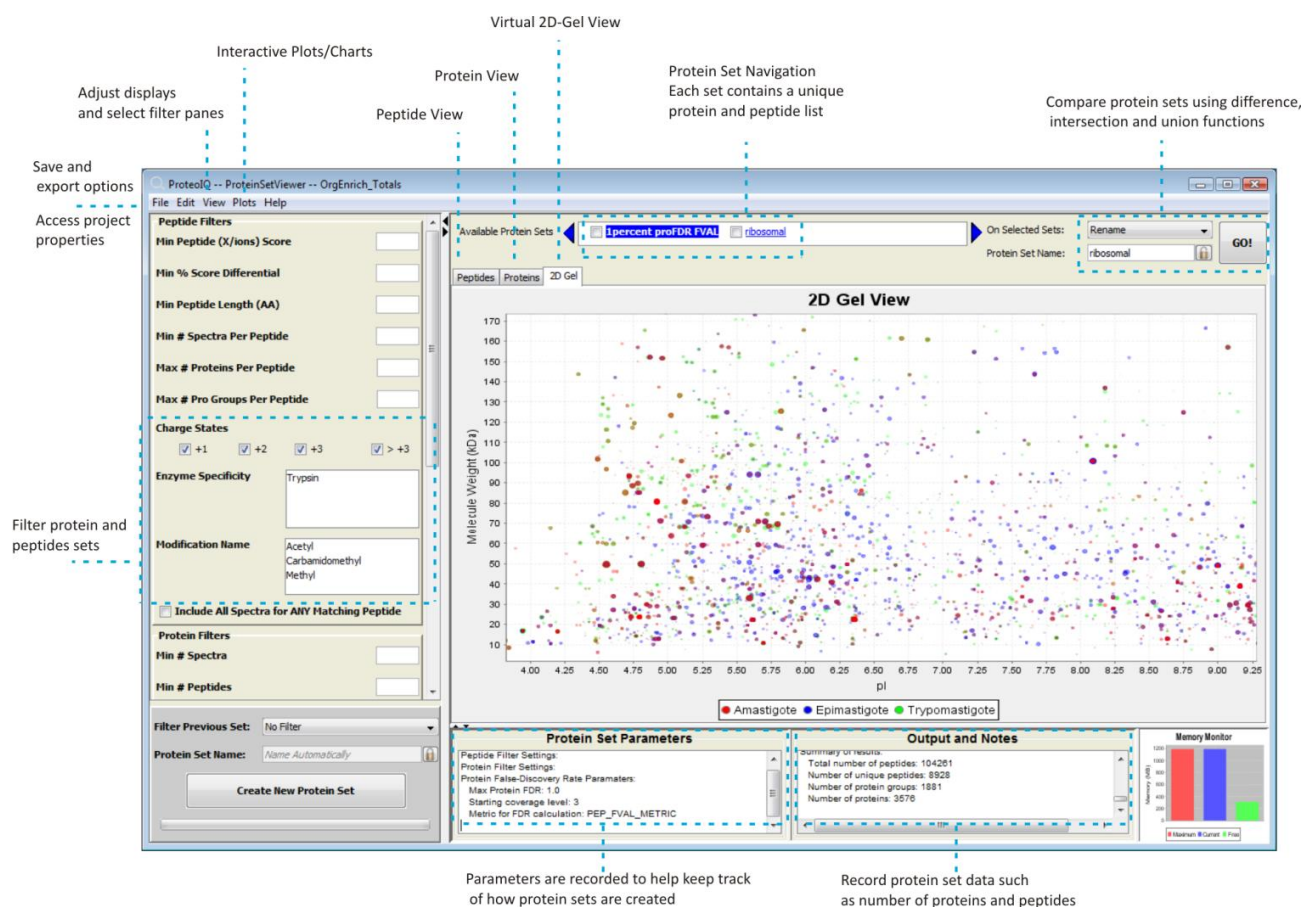
Note: For more information on adding custom hyperlinks to your database please see the “Create New Hyperlinks” tutorial.

Peptide/Protein Viewer

Figure 5 is an overview of the Protein/Peptide Set Viewer (PSV). The PSV contains all of the protein and peptide identifications in the ProteoIQ project. **Filters**, **Plots** and **Comparative Functions** are all accessible from the PSV interface. The following items can be accessed through the PSV.

- 🔍 Protein Set Navigation Bar
- 🔍 Protein Set Viewer
- 🔍 Protein Sequence Viewer
- 🔍 Peptide Set Viewer
- 🔍 MS/MS Spectra Viewer
- 🔍 Protein Set Parameters Field
- 🔍 Output and Note Field
- 🔍 Menu Bar

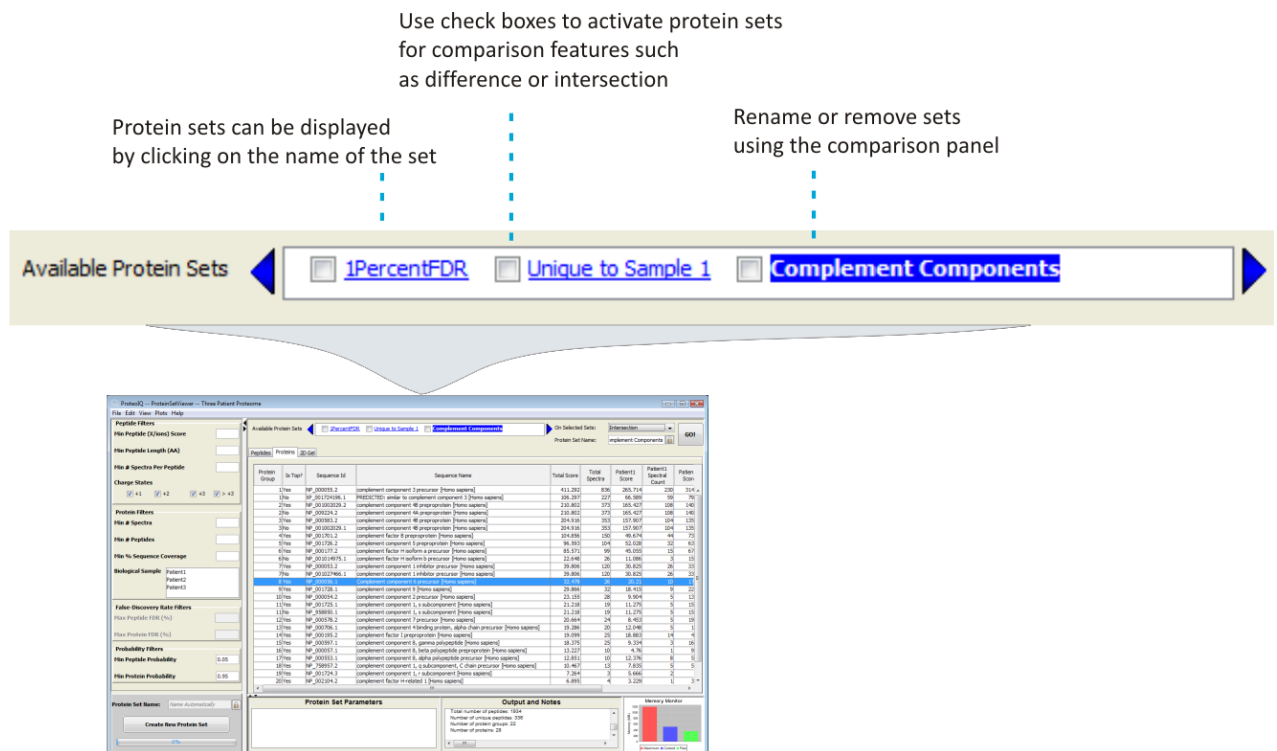
Figure 5. Protein/Peptide Set Viewer



Protein Set Navigation Bar

The protein set navigation bar is accessible from the top of the **Protein/Peptide Set Viewer** as shown in Figure 6. Each protein set is listed in the Navigation Bar and the individual protein sets are displayed by clicking on the protein set name. To compare two or more protein sets, simply click on the check box located to the left of the protein set name then select the **Comparison Function**.

Figure 6. Protein Set Navigation Bar



Protein Set Viewer

Figure 7 shows the protein set viewer. Identified proteins are organized by **Protein Group** and all relevant data such as **Score**, **Spectral Count**, and **Expression Ratios** can be accessed within the protein set viewer. Create custom views using the **Protein Table Options**.

Figure 7. Protein Set Navigation Bar

Select proteins tab to display identified proteins

All columns can be sorted or moved

Total across all biological groups

Individual data for each biological group

Protein Group	Is Top?	Sequence Id	Sequence Name	Total Score	Total Spectra	Patient1 Score	Patient1 Spectral Count	Patient1 N-SC Log2 Relative Expression
1	Yes	NP_000375.2	apolipoprotein B precursor [Homo sapiens]	434.631	496	253.063	129	-0.552
2	Yes	NP_000055.2	complement component 3 precursor [Homo sapiens]	411.292	836	265.714	230	-0.478
2	No	XP_001719515.1	PREDICTED: hypothetical protein, partial [Homo sapiens]	297.244	587	191.374	164	-0.458
2	No	XP_001724196.1	PREDICTED: similar to complement component 3 [Homo sapiens]	106.297	227	66.589	59	-0.558
3	Yes	NP_001002029.2	complement component 4B preproprotein [Homo sapiens]	210.802	373	165.427	108	-0.386
3	No	NP_009224.2	complement component 4A preproprotein [Homo sapiens]	210.802	373	165.427	108	-0.386
4	Yes	NP_000583.2	complement component 4B preproprotein [Homo sapiens]	204.916	353	157.907	104	-0.358
4	No	NP_001002029.1	complement component 4B preproprotein [Homo sapiens]	204.916	353	157.907	104	-0.358
5	Yes	NP_000087.1	ceruloplasmin precursor [Homo sapiens]	152.203	179	98.085	52	-0.375
6	Yes	NP_001701.2	complement factor B preproprotein [Homo sapiens]	104.856	150	49.674	44	-0.371
7	Yes	NP_000005.2	alpha-2-macroglobulin precursor [Homo sapiens]	99.673	128	41.95	24	-1.046
8	Yes	NP_001726.2	complement component 5 preproprotein [Homo sapiens]	96.593	104	52.028	32	-0.353
9	Yes	NP_000574.2	vitamin D-binding protein precursor [Homo sapiens]	88.866	120	45.9	28	-0.737
10	Yes	NP_000177.2	complement factor H isoform a precursor [Homo sapiens]	85.571	99	45.055	15	-1.342
10	No	NP_001014975.1	complement factor H isoform b precursor [Homo sapiens]	22.648	26	11.086	3	-1.038
11	Yes	NP_000604.1	hemopexin [Homo sapiens]	82.691	195	61.647	53	-0.481
12	Yes	NP_596869.3	titin isoform N2-A [Homo sapiens]	79.406	92	27.172	24	-0.604
13	Yes	NP_002209.2	inter-alpha (globulin) inhibitor H4 [Homo sapiens]	73.742	84	54.566	29	-0.141
14	Yes	NP_002207.2	inter-alpha (globulin) inhibitor H2 polypeptide [Homo sapiens]	65.591	86	37.901	26	-0.443
15	Yes	NP_001076.2	serpin peptidase inhibitor, clade A, member 3 precursor [Homo sapiens]	65.531	176	48.367	60	-0.201
16	Yes	NP_000884.1	kininogen 1 isoform 2 [Homo sapiens]	62.489	104	34.879	25	-0.677
16	No	NP_001095886.1	kininogen 1 isoform 1 [Homo sapiens]	55.964	87	28.353	17	-1.038
17	Yes	NP_570602.2	alpha 1B-glycoprotein precursor [Homo sapiens]	60.067	77	28.532	12	-1.386
18	Yes	NP_000497.1	coagulation factor II precursor [Homo sapiens]	59.524	60	28.717	15	-0.656
19	Yes	NP_000473.2	apolipoprotein A-IV precursor [Homo sapiens]	58.287	81	32.426	16	-0.964
20	Yes	NP_002206.2	inter-alpha (globulin) inhibitor H1 [Homo sapiens]	57.771	69	31.107	11	-1.429

One protein or multiple proteins can be selected by a single right click on the row of interest (Figure 8). Right clicking on the selected proteins allows the highlighted proteins to be extracted to a new protein set that is accessed via the Protein Set Navigation Bar. Selecting View Protein Info opens the **Protein Sequence Viewer** as shown in Figure 9.

Figure 8. Right click functionality in the Protein Set Viewer

Single clicking on rows highlights/selects the proteins of interest

Protein Group	Is Top?	Sequence Id	Sequence Name
1	Yes	NP_000375.2	apolipoprotein B precursor [Homo sapiens]
2	Yes	NP_000055.2	complement component 3 precursor [Homo sapiens]
2	No	XP_001719515.1	PREDICTED: hypothetical protein, partial [Homo sapiens]
2	No	XP_001724196.1	PREDICTED: similar to complement component 3 [Homo sapiens]
3	Yes	NP_001002029.2	complement component 4B preproprotein [Homo sapiens]
3	No	NP_009224.2	complement component 4A preproprotein [Homo sapiens]
4	Yes	NP_000583.2	complement component 4B preproprotein [Homo sapiens]

Copy Selected (Ctrl+C)

Copy Selected to New Protein Set

View Protein Info

View Protein Info opens the Proteins Sequence Viewer

Copy Selected to New Protein Set creates a new protein set that containing only the selected proteins

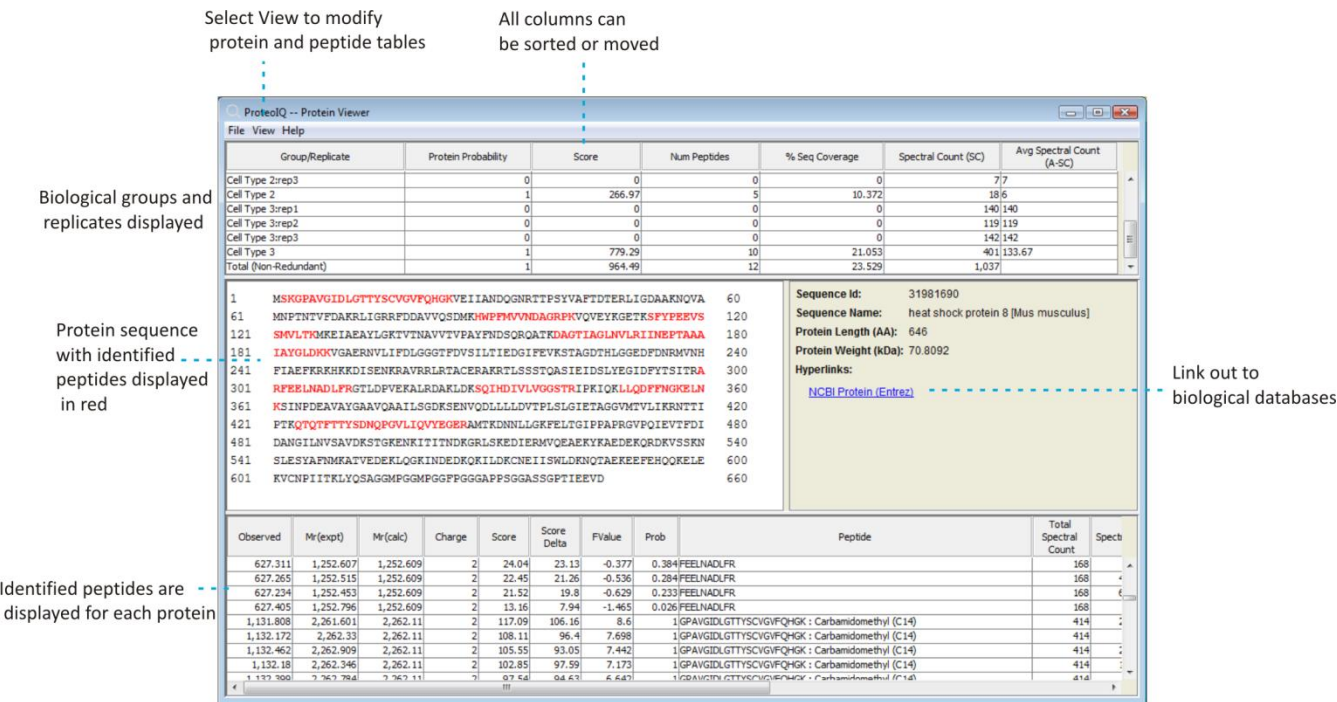
Protein Sequence Viewer

Detailed information about every identified protein can be accessed in the Protein Sequence Viewer shown in Figure 9. The Protein Sequence Viewer is opened by double clicking on the protein name in the [Protein Set Viewer](#).

Identified peptides are highlighted in red and overlaid on the full amino acid sequence for the protein.

The top table contains information about the protein identification broken down by biological group and replicate.

Figure 9. Protein Sequence Viewer

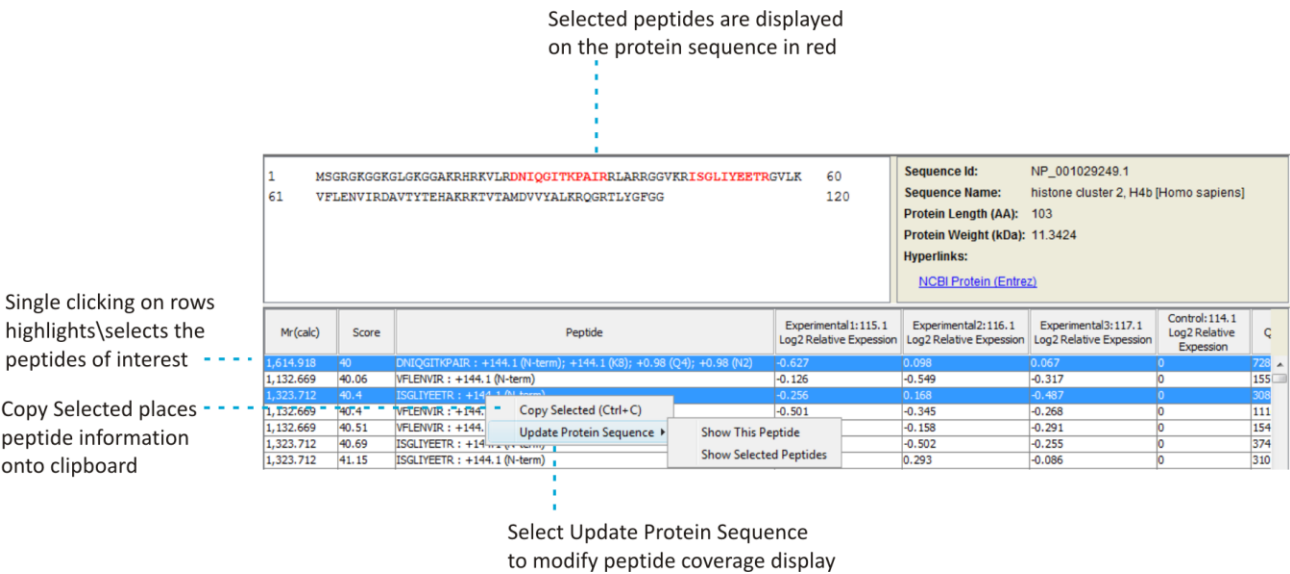


The lower table provides information about each peptide identification. Double clicking on the peptide sequence will access the [MS/MS Spectral Viewer](#). Right clicking on the peptide sequence opens up the Update Protein Sequence dialogue box.

Hyperlinks created in the [Modify Database Wizard](#) are displayed in the right panel.

One or more peptides can be selected by a single right click on the row of interest (Figure 10). Right clicking on the selected peptides allows the highlighted peptides to be displayed in the protein sequence coverage. The selected peptides are displayed on the protein sequence in red. Selecting Copy Selected places the peptide information on the clipboard.

Figure 10. Right click functionality in the Protein Sequence Viewer



Peptide Set Viewer

The Peptide Set Viewer displays the identified peptides (Figure 11). Information about the peptide identifications are displayed along with each peptide sequence and any PTM's predicted to be present on the peptide are listed under the modification column. Customize the peptide table by selecting [Peptide Table Options](#).

Figure 11. Peptide Set Viewer

Select peptides tab to display identified peptides

All columns can be sorted or moved

Double click on peptide sequence to access MS/MS spectral view

Find peptides with PTMs by sorting the modification column

Mr(expt)	Mr(calc)	Charge	Score	Delta	Prob	Peptide Sequence	Modification	Total Spectral Count	Spectra Intensity
1,805.976	1,805.962	2	5.028	2.833	0.951	1FIQGVISWGVVDVCK	+57.0215 (C15)	3	11,102.4
3,343.866	3,343.782	4	2.143	1.134	0.951	FILSIATSPDGKYLASGAIDGIINFDIATGK		1	7,423.7
1,079.594	1,079.573	2	3	1.062	0.98	NYNLVESLK		11	5,895.4
1,773.018	1,772.977	2	4.563	2.655	1	FICPLTGLWPINTLK	+57.0215 (C3)	9	6,408
1,121.608	1,121.584	2	3.242	1.414	0.995	FGLDDEGK		11	7,628.4
1,274.614	1,274.585	2	3.016	0.843	0.924	NWGLGHAFCR	+57.0215 (C10)	1	4,862.2
1,052.492	1,052.487	2	2.418	1.038	0.979	WGYLEPK	+57.0215 (C4)	2	4,098.2
1,180.606	1,180.582	2	1.941	0.086	0.15	WGYLEPK	+57.0215 (C4)	2	8,251.5
2,835.388	2,835.414	3	2.89	1.283	0.959	FFLTCVSGEAGAGRSDAIWGPILLEK	+57.0215 (C5)	2	4,759
3,683.842	3,683.752	4	5.089	3.307	0.988	NVQNYPHSTVDVTQNNFHYFGSSEIVVAGK		3	5,641.7
2,631.394	2,631.433	3	2.902	1.139	0.956	FFGLFYAVGTSPNLPLYTLRINK		2	9,385.9
1,831.023	1,830.986	2	1.78	0.078	0.068	NVAGTRSVAVNCKVLDK	+57.0215 (C12)	1	8,252.3
1,046.598	1,046.588	2	3.341	1.199	0.99	FEVQIVYK		9	7,085.2
2,129.142	2,129.122	3	2.194	0.403	0.378	WFLLEQETQVAHFFPK		1	5,184.9
1,684.526	1,684.892	3	1.641	0.466	0.488	POPVRGEVPPRTPR		26	2,689.3
1,268.703	1,268.675	2	2.01	0.701	0.665	POPVRGEVPPR		26	3,663.9
1,140.574	1,140.551	2	2.575	1.079	0.984	WEVYCLNK	+57.0215 (C4)	1	4,579.5
3,985.35	3,985.254	3	1.212	1.097	1	NTRLGWIQKQVTVLGGPVPVWFLGWFAAPPLGSLR		1	7,899.3
1,909.801	1,909.77	3	1.392	0.335	0.168	WESCDYQRSECHPR	+57.0215 (C4); +57.0215 (C11)	1	5,470.9
1,092.651	1,092.63	2	2.956	1.207	0.99	NLIIVLDK		11	4,364.2
2,481.295	2,481.233	3	5.278	2.829	0.999	NLIIVLDKVSSEDDCLAFK	+57.0215 (C17)	8	5,632.2
1,508.751	1,508.741	2	2.052	0.039	0.093	WEPOSQKGMTVK	+15.9949 (M10)	1	8,081.5
2,299.26	2,299.204	3	3.625	0.785	0.79	FDASHKIEVEGTRGAVELNK		1	9,355.6
1,370.668	1,370.637	2	3.384	1.894	0.999	NFTLSCDGLSR	+57.0215 (C7)	3	4,143.8
1,088.536	1,088.519	2	3.153	1.2	0.993	WELCDIPR	+57.0215 (C4)	12	7,268.5
2,096.064	2,096.014	3	3.826	1.76	0.995	NTEOEAGAEVHEVEYDK		3	8,037.3

One peptide or multiple peptides can be selected by a single right click on the row of interest (Figure 12). Right clicking on the selected peptides allows the highlighted peptides to be extracted to a new protein set that is accessed via the [Protein Set Navigation Bar](#). Select Matching Proteins highlights proteins containing the selected peptides in the Protein Set Viewer.

Figure 12. Right click functionality in the Peptide Set Viewer

Single clicking on rows highlights\selects the peptides of interest

Select Matching Proteins highlights the proteins in the Protein Set Viewer that contain the selected peptides

Copy Selected to New Protein Set creates a protein set ONLY containing the selected peptides

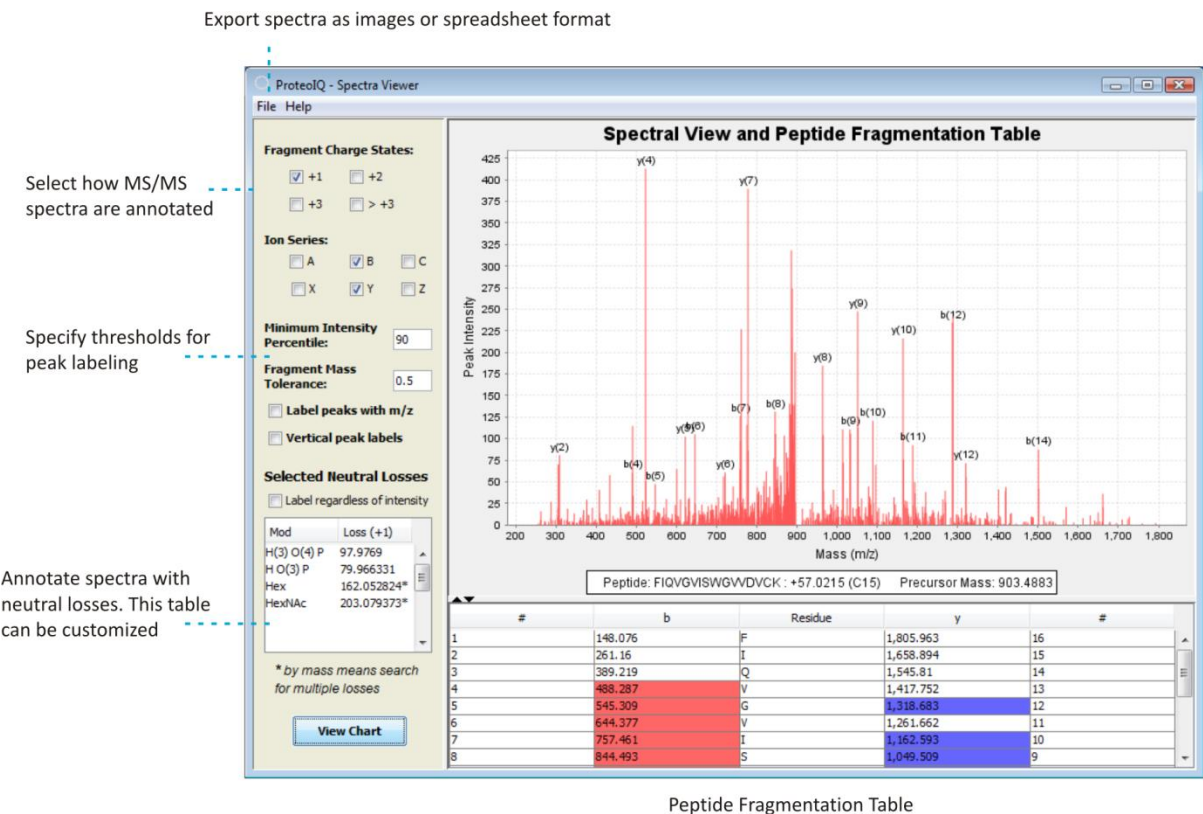
Mr(expt)	Mr(calc)	Charge	Score	Delta	Prob	Peptide Sequence	Modification
1,805.976	1,805.962	2	5.028	2.833		1FIQGVISWGVVDVCK	+57.0215 (C15)
3,343.866	3,343.782	4	2.143	1.134	0.951	FILSIATSPDGKYLASGAIDGIINFDIATGK	
1,079.594	1,079.573	2	3	1.062	0.98	NYNLVESLK	
1,773.018	1,772.977	2	4.563	2.655	1	FICPLTGLWPINTLK	+57.0215 (C3)
1,121.608	1,121.584	2	3.242	1.414	0.995	FGLDDEGK	
1,274.614	1,274.585	2	3.016	0.843	0.924	NWGLGHAFCR	+57.0215 (C10)
1,052.492	1,052.487	2	2.418	1.038	0.979	WGYLEPK	+57.0215 (C4)
1,180.606	1,180.582	2	1.941	0.086	0.15	WGYLEPK	+57.0215 (C4)
2,835.388	2,835.414	3	2.89	1.283	0.959	FFLTCV	+57.0215 (C5)

MS/MS Spectra Viewer

Figure 13 is an example of the MS/MS Spectra Viewer. To open the spectra viewer, double click on the peptide sequence in the [Peptide Set Viewer](#) or in the [Protein Sequence Viewer](#).

The spectra viewer contains three panels. The left most panel is for selection of spectra labeling parameters that are rendered in the annotated MS/MS spectrum. Select to annotate spectra by [Charge State](#), [Fragment Ion Type](#), or search for [Neutral Losses](#).

Figure 13. MS/MS Spectra Viewer



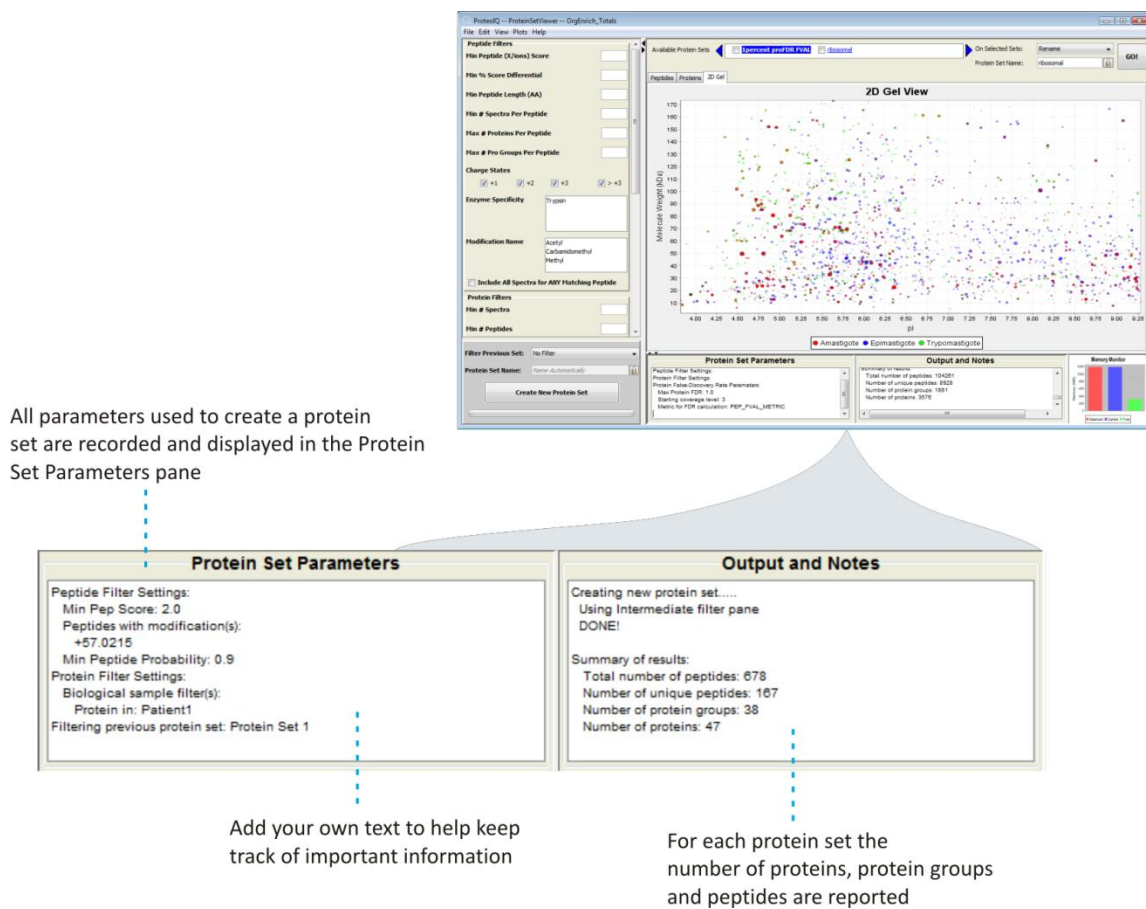
The bottom panel contains a table with the predicted m/z of the selected ion series. The color codes indicate that the specified ion series was detected in the selected MS/MS spectrum.

Protein Set Parameters Field and Output Notes

Since multiple Protein Sets can be created in ProteoIQ, the Parameters and protein set information are reported in the Protein Set Parameters and Output Notes panes (Figure 14). Both are located at the bottom of the Protein/Peptide Set Viewer.

Custom text may be entered into these text fields to help keep track of important information related to each Protein Set.

Figure 14. Protein Set Parameters and Output Notes



Menu Bar

The following Menus can be accessed via drop down boxes from the Menu Bar and by right clicking in the Protein/Peptide Set Viewer and Plots (Pop-Up Menus).

- 🔍 [File Menu](#)
- 🔍 [Edit Menu](#)
- 🔍 [View Menu](#)
- 🔍 [Plots Menu](#)
- 🔍 [Help Menu](#)
- 🔍 [Pop-Up Menus](#)

File Menu

The File Menu is used to access [Save](#) and [Export](#) functions or to review Project Properties (Figure 15). Items in the File Menu are described in Table 1.

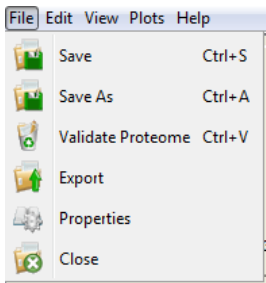


Figure 15. File Menu

Table 1. File Menu Items

Menu Item	About	Access	Short Cut Key
Save	Saves ProteoIQ project	File > Save	Ctrl +S
Save As	Save ProteoIQ project	File > Save As	Ctrl +A
Validate Proteome	Save Validated Proteome	File > Validate Proteome	Ctrl +V
Export	Export proteins, peptides, and spectra	File > Export	
Properties	Access Project Properties	File > Project Properties	
Close	Close program	File > Close	

Edit Menu

The Edit Menu (Figure 16) accesses the Find, Copy and Edit functions as described in

Table 2

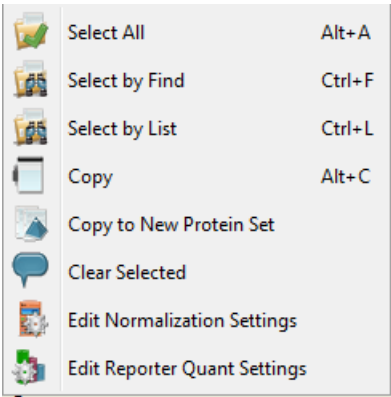


Figure 16. Edit Menu

Table 2. Edit Menu Items

Menu Item	About	Access	Short Cut Key
Select All	Selects all proteins or peptides in the peptide or protein set viewer	Edit > Select All	Alt +A
Select by Find	Selects data in the peptide or protein set viewer based on user defined search criteria	Edit > Select by Find	Ctrl +F
Select by List	Load a tab-delimited file containing user defined search criteria and selects data in the peptide or protein set viewer	Edit > Select by List	Ctrl + L
Copy	Copy selected information	Edit > Copy	Alt +C
Copy to New Protein Set	Copies selected peptides or proteins to a new Protein Set accessible in the Navigation Bar	Edit > Copy to New Protein Set	
Clear Selected	Deselects selected items in peptide and protein set viewer	Edit > Clear Selected	

Menu Item	About	Access
Edit Normalization Settings	Opens the Normalization Settings dialogue box	Edit > Edit Normalization Settings
Edit Reporter Quant or Label Free Settings	Opens the edit Reporter Quantitation or Label Free Settings dialogue box	Edit > Edit Reporter Quant/Label Free Settings

View Menu

From the View Menu (Figure 17) the **Filter Panes**, **Peptide** and **Protein Table Options** are accessible. Table 3 lists the menu items availabe in the View Menu.

Figure 17. View Menu

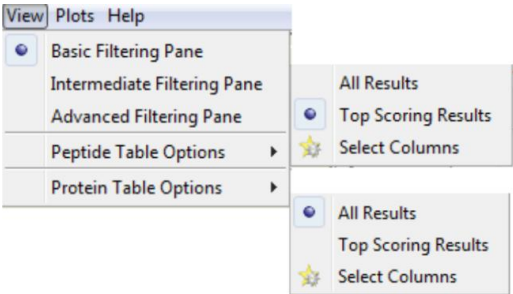


Table 3. View Menu Items

Menu Item	About	Access
Basic Filtering Pane	Toggle Basic Filtering Pane	View > Basic Filtering Pane
Intermediate Filtering Pane	Toggle Intermediate Filtering Pane	View > Intermediate Filtering Pane
Advanced Filtering Pane	Toggle Advanced Filtering Pane	View > Advanced Filtering Pane
Peptide Table Options	Adjust display options in the Peptide Set Viewer	View > Peptide Table Options
Protein Table Options	Adjust display options in the Protein Set Viewer	View > Protein Table Options

Plots Menu

Graphical displays are available in the Plots Menu (Figure 18). Table 4 lists the menu items available in the Plots Menu.

Figure 18. Plots Menu

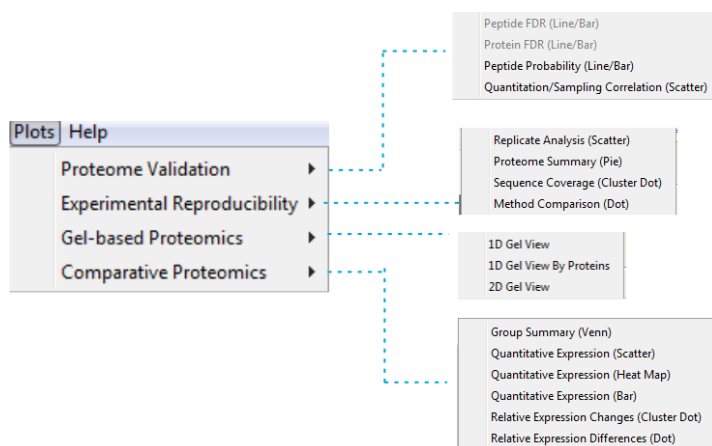


Table 4. Plots Menu Items

Menu Item	About	Access
Peptide FDR	Opens the Peptide FDR Plot	Plots > Proteome Validation > Peptide FDR
Protein FDR	Opens the Protein FDR Plot	Plots > Proteome Validation > Protein FDR
Peptide Probability	Opens the Peptide Probability Plot	Plots > Proteome Validation > Peptide Probability
Quantitation/Sampling Correlation (Scatter)	Opens the Quantitation/Sampling (Scatter) Plot	Plots > Proteome Validation > Quantitation/Sampling Correlation (Scatter)
Replicate Analysis (Scatter)	Opens the Replicate Analysis (Scatter) Plot	Plots > Experimental Reproducibility > Replicate Analysis (Scatter)
Proteome Summary (Pie)	Opens the Pie Chart	Plots > Experimental Reproducibility > Proteome Summary (Pie)
Sequence Coverage (Cluster Dot)	Opens the Sequence Coverage (Cluster Dot) Plot	Plots > Experimental Reproducibility > Sequence Coverage (Cluster Dot)

Menu Item	About	Access
Method Comparison	Opens the Method Comparison Plot	Plots > Experimental Reproducibility > Method Comparison (Dot)
1D Gel View	Opens the 1D Gel View	Plots > Gel-based Proteomics > 1D Gel View
1D Gel View by Protein	Opens the 1D Gel View by Protein	Plots > Gel-based Proteomics > 1D Gel View by Protein
2D Gel View	Opens the 2D Gel View	Plots > Gel-based Proteomics > 2D Gel View
Group Summary (Venn Diagram)	Opens the Venn Diagram	Plots > Comparative Proteomics > Venn Diagram
Quantitative Expression (Scatter)	Opens the Quantitative Expression (Scatter) Plot	Plots > Comparative Proteomics > Quantitative Expression (Scatter)
Quantitative Expression (Heat Map)	Opens the Quantitative Expression (Heat Map) Plot	Plots > Comparative Proteomics > Quantitative Expression (Heat Map)
Quantitative Expression (Bar Chart)	Opens the Quantitative Expression (Bar Chart) Plot	Plots > Comparative Proteomics > Quantitative Expression (Bar Chart)
Relative Expression Changes (Cluster Dot)	Opens the Relative Expression Changes (Cluster Dot) Plot	Plots > Comparative Proteomics > Relative Expression Changes (Cluster Dot)
Relative Expression Differences (Dot)	Opens the Relative Expression Differences (Dot) Plot	Plots > Comparative Proteomics > Relative Expression Differences (Dot)

Help Menu

Select the help menu to access the onboard ProteoIQ User Guide and to view software versions (Figure 19)

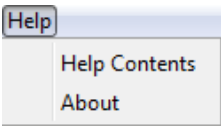


Figure 19. Help Menu

Table 5. Help Menu

Menu Item	About	Access
Help Contents	Opens onboard help guide	Help > Help Contents
About	Opens the About page to display version information	Help > About

Pop-Up-Menus

Right clicking on the selected peptides or proteins in the Peptide and Protein Set Viewer opens the Pop-Up-Menus as shown in Figure 12 and Figure 8. Pop-Up-Menu items are listed in Table 6.

Table 6. Pop-Up-Menu

Menu Item	About	Access	Shortcut
Copy	Copy selected information	Right click > Copy Selected	Ctrl +C
Copy to New Protein Set	Copies selected peptides or proteins to a new Protein Set accessible in the Navigation Bar	Right click > Copy to New Protein Set	
Select Matching Proteins	Selects proteins in Protein Set Viewer that contain selected peptides	Right click > Select Matching Proteins	
View Protein Info	Opens the Protein Sequence Viewer	Right click > View Protein Info	
Normalize SC from selected proteins	Opens the Normalize SC From Selected Proteins dialogue box.	Right click > Normalize SC from selected proteins	

Statistical Algorithms Overview

ProteoIQ provides two methods for statistical analysis of peptide and protein identifications. Determining confidence in peptide and protein assignments can be performed using [Peptide Probability](#) or [Protein Probability](#) thresholds or [False Discovery Rate](#) analysis. See the discussion on "[Evaluating Confidence in Peptide and Protein Identifications](#)" for more details on how to apply probabilities or false discovery rates in ProteoIQ.

Probability

ProteoIQ calculates probabilities that peptide and protein assignments are correct by employing a Java based implementation of the statistical models commonly known as Peptide Prophet and Protein Prophet, as developed by the Institute for Systems Biology. In the Peptide Set Viewer, probability refers to the probability that the peptide assignment to a given spectra is correct. In the Protein Set Viewer, probability indicates the likelihood that the protein assignment is correct taking into account the probability of the peptide assignments for all peptides apportioned to that protein. For a more detailed discussion on how the probabilities are calculated see the journal articles below.

References:

- Keller, A., Nesvizhskii, A. I., Kolker, E., Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 2002, 74, 5383-5392.
- Nesvizhskii, A. I., Keller, A., Kolker, E., Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 2003, 75, 4646-4658.

Peptide and Protein False Discovery Rates

ProteoIQ utilizes the ProValT algorithm to calculate protein and peptide false discovery rates as described in Weatherly *et al.* False discovery rates are calculated by comparing the frequency of peptide and protein identifications at score thresholds between a Target and Decoy (reversed or scrambled) database. The false discovery rate is defined as the number of decoy peptide or protein identifications divided by the number of target peptides or proteins multiplied by 100%.

Reference:

- Weatherly, DB., Atwood, JA., Minning, TA., Cavola, C., Tarleton, RL., Orlando, R. A heuristic method for assigning a false-discovery rate for protein identifications from Mascot database search results. *Mol.*

Data File Formats

Table 7 describes the files formats supported in ProteoIQ.

Table 7. File formats

File Extension	About	Location and Use
.dat	Mascot (Matrix Science) database search result format.	Import in New Project Wizard
.out	Thermo Fischer database search result format containing peptide information	Import in New Project Wizard
.dta	Thermo Fischer MS/MS peak list format used to create the MS/MS spectral view in ProteoIQ	Import in New Project Wizard
.srf	Thermo Fischer, BioWorks search results file.	Import in New Project Wizard
.piq	ProteoIQ project extension	Saved when a project is created Select this file to open a project
.fasta	Database format	Export
.obj	ProteoIQ project file	Required to be present in project folder but not selected to open a project
.inx	ProteoIQ index file	Required to be present in project folder but not selected to open a project
.xml	X!Tandem database search results output format	Import in New Project Wizard

How to Use ProteoIQ

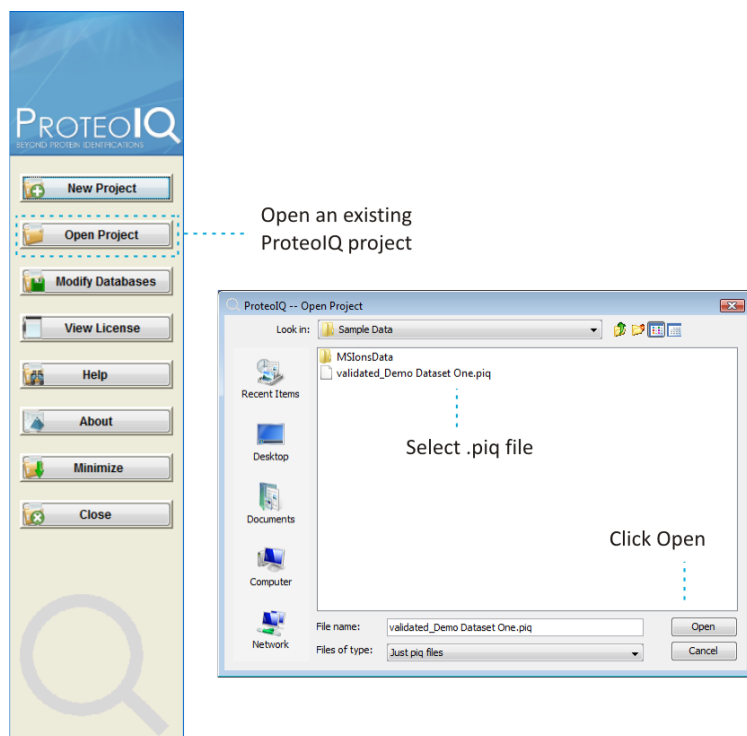
The following sections provide an overview about how to use ProteoIQ. The following contents are covered in this section.

- 🔍 [Opening a ProteoIQ Project](#)
- 🔍 [Saving Projects](#)
- 🔍 [Sharing/Moving ProteoIQ Projects](#)
- 🔍 [Creating a New ProteoIQ Project Overview](#)
- 🔍 [What You Need to Create a New Project](#)
- 🔍 [Creating a New ProteoIQ Project – Label Free Quantitation](#)
- 🔍 [Creating a New ProteoIQ Project – Reporter Ion Quantitation](#)
- 🔍 [Modifying Databases](#)
- 🔍 [Using the Filter Pane](#)
- 🔍 [Comparing Protein Sets](#)
- 🔍 [Customizing Results Views](#)
- 🔍 [Using the Plots](#)
- 🔍 [Exporting Results](#)

Opening a ProteoIQ Project

This section describes how to open a ProteoIQ Project and provides instructions for accessing the Demo ProteoIQ Project as shown in Figure 20.

Figure 20. Open Project



Open a ProteoIQ Project

An Demo ProteoIQ Project is including with the installation package.

To open the Demo ProteoIQ Project:

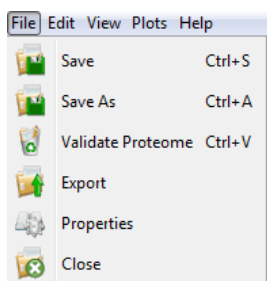
1. Select **Navigation Menu > Open Project**
2. Browse to **C:\Program Files\ProteoIQ\Sample Data\validated_Demo_Dataset_One.piq**
3. Select Open

Note: A ProteoIQ Project contains multiple files. For the project to opened correctly ALL files must be present in the source folder. Select only the .piq file to open a ProteoIQ project.

Saving a ProteoIQ Project

There are two options for saving a ProteoIQ Project, **Save** and **Validate Proteome**. The following sections describe both saving options and discuss when each should be applied. For a detailed discussion on the difference between Save and Validate Proteome see the “**What is the difference between Save and Validate**” discussion section. Figure 21 shows how to access the Save functions. Table 8 provides an overview of the Save functions.

Figure 21. Project Save



To Save a ProteoIQ Project:

1. Select **File > Save or Save As**
2. Browse to directory for file to be saved
3. Select Save

Note: The Save or Save As functions create a file containing ALL of the peptide and proteins identifications in the ProteoIQ Project even if the peptides or proteins are not displayed in the Peptide/Protein Set Viewer. See the “What is the difference between Save and Validate” discussion section for additional information.

To Validated a ProteoIQ Project:

1. Select **File > Validate Proteome**
2. Browse to directory for file to be saved
3. Select Save

Note: Validate Proteome DISCARDS ALL decoy and target peptides/protein/spectra that are not displayed in the Peptide or Protein Set Viewer. It is recommended that you Validate prior to sharing data with collaborators since the resulting files sizes are significantly smaller. See the “What is the difference between Save and Validate” discussion section for additional information.

What is the Difference Between Save and Validate?

When you Save a ProteoIQ Project you are saving a Whole Project. The key difference between Whole Project and Validated is in the spectra/peptides that are saved. The Whole Project data will contain **ALL** spectra/peptides that were parsed for both target and decoy data. This way you can always open the Whole Project and perform validation on your proteome. When you do a validation, however, you are discarding all decoy data and all target spectra/peptides that are not used in the protein sets that are shown at the time that you validate.

When you first create a new project, you are building the Whole Project data set. The spectra, peptides, peptide to protein mappings, etc are saved automatically after parsing and clustering in the Whole Project folder. Then, the Protein Set Viewer is opened and the initial protein set is generated. Again, this protein set has no bearing, necessarily; on other protein sets created on the Whole Project (e.g. you can do an initial protein set at 1% FDR and then do subsequent ones at 5% or 0.5% using all peptides parsed from target/decoy results. However, the initial protein set (and subsequent ones) are not automatically saved for you. You have to manually Save these.

Validating a ProteoIQ Project

A Validated ProteoIQ Project discards all decoy data and all target spectra/peptides that are not used in the protein sets that are shown at the time that you validate. Validated projects contain only peptides and proteins viewable in the protein set viewer. You **CAN** do further validation, but you will never be able to add data to the validated proteome.

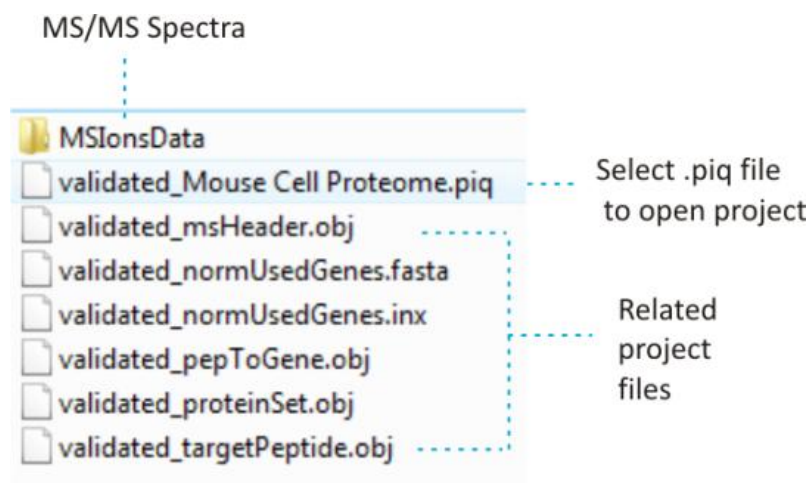
Table 8. Summary of Save Features

Save Function	Description
Save	Saves the current protein sets shown in the protein set viewer and all unused peptides so that validation can be performed at a later date. All data is saved in the project path directory specified in the New Project Wizard.
Save As	Save As will save all data, depending on if you are currently working on the Whole Project or Validated project, in another folder.
Validate Proteome	Saving a Validate Proteome removes all unused peptides/spectra from the dataset to save memory and then issues a Save As. Validated proteomes are small files that are excellent for sharing with collaborators.

Sharing and Moving a ProteoIQ Project

Figure 22 shows a typical ProteoIQ Project results folder. Each folder contains the .piq file which should be used to [Open a ProteoIQ Project](#). Index, object and database files are also present. To share or move a ProteoIQ Project, the ENTIRE folder must be transferred.

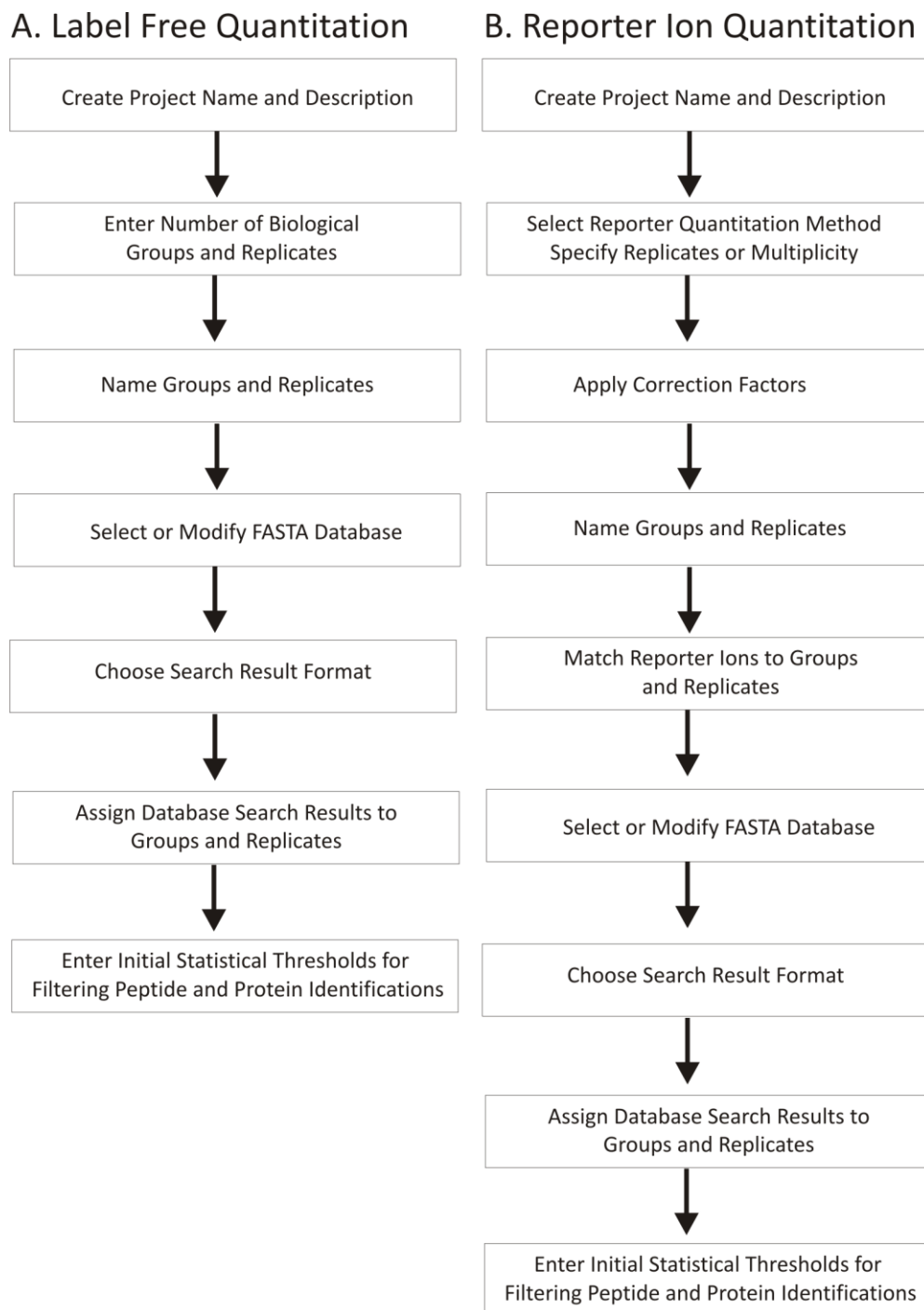
Figure 22. A typical ProteoIQ Project Folder



Creating a New Project Overview

Figure 23 shows the process taken to create a new ProteoIQ Project. The New Project Wizard and [Modify Database Wizard](#) provide step-by-step instructions for creating a new project.

Figure 23. The New Project Process for Label Free Quantitation (A) and Reporter Ion Quantitation (B)



What You Need to Create a New Project

Table 9 describes the files required to **Create a New Project** in ProteoIQ. For more details on creating new projects in ProteoIQ see the “Database Wizard” and “Grouping Database Search Results” Tutorials located at **C:\Program Files\ProteoIQ\docs**.

Table 9. Files required to create a new ProteoIQ project

File Type	Description	File Extension
Database Search Result (Mascot)	Contains search parameters, peptides, protein assignments, and MS/MS spectra.	.DAT
Database Search Result (XI!Tandem)	Contains search parameters, peptides, protein assignments, and MS/MS spectra.	.XML
Database Search Result (SEQUEST).	SEQUEST creates folders containing peptide files (.out), spectral files (.dta), and parameters files. ProteoIQ requires all three.	.OUT, .DTA, and .PARAMS
	.SRF files can also be imported into ProteoIQ	or .SRF
FASTA Database	Contains the protein sequences used to perform the database search.	.FASTA

Mascot (.DAT)

Mascot renders results in two formats: its native data format (.DAT) and an HTML view. The HTML files should NOT be used in ProteoIQ since they do not contain all of the information needed to perform protein grouping and statistical analysis. The .DAT files are typically located in the directory Inetpub\Mascot\data and are organized by an assigned number which is generated when the search is performed.

Often determining which .DAT file corresponds to a particular peak list may be difficult; therefore, we recommend copying the search results into another directory prior to running ProteoIQ. This will ensure that the data may be easily accessed for multiple analyses at a later date. There is also no need to rename your .DAT files. When you select the file in the Available Results Files Viewer, ProteoIQ extracts the name of the peak list from the .DAT file and renders it in the window, thus making it easy to determine which database results correspond to each experiment.

SEQUEST (.OUT, .DTA, .PARAMS and .SRF)

ProteoIQ supports either binary search results formats (.SRF) or the human readable format of SEQUEST .OUT, .DTA, and .PARAMS. For the .OUT format, each search result folder should contain the .OUT, .DTA, and .PARAMS files.

Note: SEQUEST folders often contain very large numbers of files. For this reason, we recommend copying these folders to a separate directory prior to running ProteoIQ.

BioWorks versions 3.2 and greater support the creation of .SRF files. ProteoIQ currently support parsing of .SRF files generated directly from the SEQUEST search or via conversion of the .OUT and .DTA results folders to .SRF format.

X!Tandem (.XML)

X!Tandem generates an .XML results report than can be directly imported into ProteoIQ.

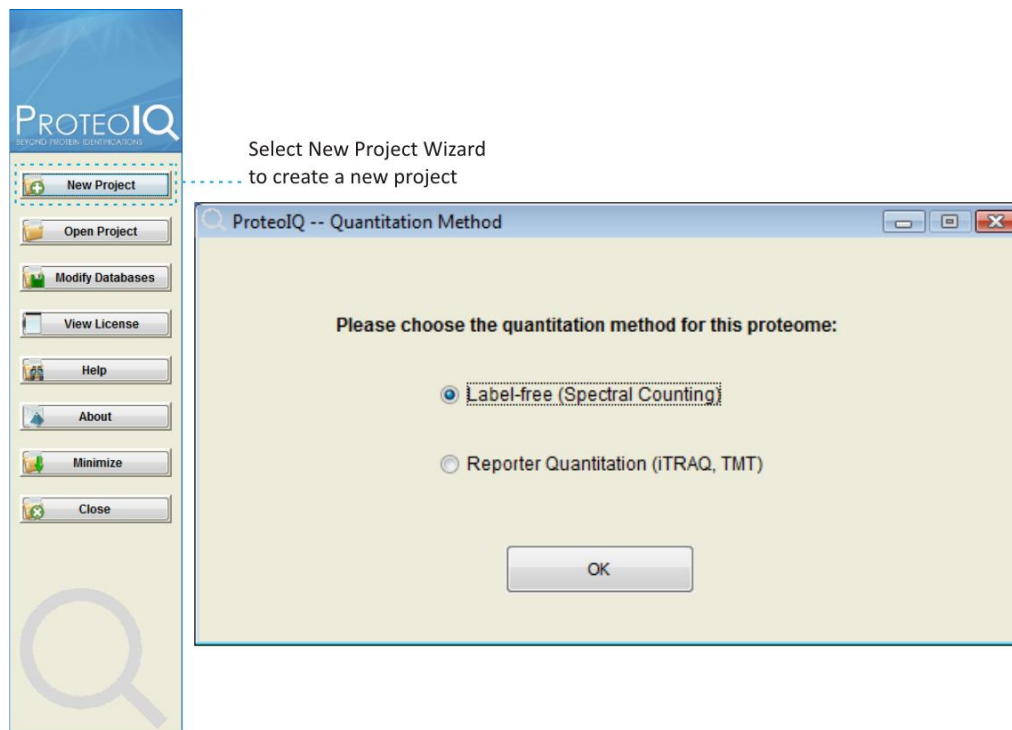
FASTA Database

ProteoIQ maps peptide identifications to proteins present in a FASTA database. To ensure accurate results, the same database used to generate the search results should be used to create the ProteoIQ project. See [Modify Database Wizard](#) for information on importing databases in ProteoIQ.

Creating a New Project – Label Free Quantitation

To create a new project in ProteoIQ, first select the New Project Wizard from the Navigation Pane as shown in Figure 24, then select Label Free Quantitation.

Figure 24. New Project Wizard



There are twelve sections in the New Project Wizard. The following describe the individual sections in the New Project Wizard.

- Q [Describe Project](#)
- Q [Describe Proteome](#)
- Q [Describe Samples](#)
- Q [Choose Database](#)
- Q [Search Results Options](#)
- Q [Select Target Files for Parsing](#)
- Q [Select Decoy Files for Parsing](#)
- Q [Select Non Target Files for Parsing](#)
- Q [Peptide Parsing Options](#)
- Q [Parse Files\MS Ions Data](#)
- Q [Cluster Peptides](#)

Describe Project

The Describe Project section contains fields for adding information about the new project and defines where the project will be saved. Figure 25 shows the Describe Project section of the **New Project Wizard**. Table 10 describes the fields.

Figure 25. Describe Project

Table 10. Fields in the Describe Project Section

Field	Description
Project Name	The project name should be entered in this field. A ProteoIQ project may contain multiple proteomes and peptide/protein sets along with all the parameters used in its creation. A project may be saved and reopened at any time.
User Name	Enter a user name here to assist you in identifying your project in the future. This field is not required.
Project Description	Enter a description of the project here to assist you in identifying your project. This field is optional.
Project Path	Specifies the location that the project will be saved. Click “browse” to locate the project path on your computer or network.

Describe Proteome

The Describe Proteome section contains fields for adding information about the project and defines how the database search results will be grouped (Figure 26). Table 11 describes the fields.

Figure 26. Describe Proteome

Table 11. Fields in the Describe Proteome Section

Field	Description
Proteome Name	Enter the name of the proteome here. A descriptive name should be given, as a single project may contain multiple proteomes.
Instrument Description	Enter the type of mass spectrometer used in the proteome analysis. This is an optional field and has no effect on the data processing.
Combine all results files (MudPIT)	Groups all database search results as if from a single biological sample. See “ Combine All Results Files ” for detailed information.
Each result represents an individual sample	Performs protein identification in each database search results using ONLY peptides identified from that specific result file. See topic “ Each Result Represents an Individual Sample ” for more information.
Group files by biological sample	Allows the user to define the number of biological samples and the number of replicate analyses performed on each sample. Results files from all replicates of the same biological sample are then combined for further analysis.

Field	Description
Spectra were searched against a decoy database	This box allows the user to specify that a decoy database search was performed. Searching against a decoy database is used when ProteoIQ calculates False Discovery Rates (FDR) for the peptides and proteins identified. Specify if the database was concatenated or separate target and decoy databases.
Spectra were search against a non-target database	Specifies if a non-target database search was performed. This option is useful if one expects that the sample contained proteins from organisms other than the target organism (e.g. human keratins in protein preparations).

Combine All Results Files (MudPIT)

Groups all database search results as if from a single biological sample. This is useful when peptides from a single biological sample are subjected to multidimensional chromatography-tandem mass spectrometry with off-line fractionation between the chromatographic steps. For example, the peptides are separated into 10 fractions by strong cation exchange chromatography, each of which is analyzed by reversed-phase chromatography- tandem mass spectrometry. This experiment would result in a series of 10 raw files, which would be processed into peak list format and searched separately generating 10 Mascot .dat or SEQUEST results folders (containing many .out files). By selecting this option, ProteoIQ assumes all of the peptides were derived from a single source - regardless of the fact that the peptides were present in different fractions and thus were identified in distinct search results.

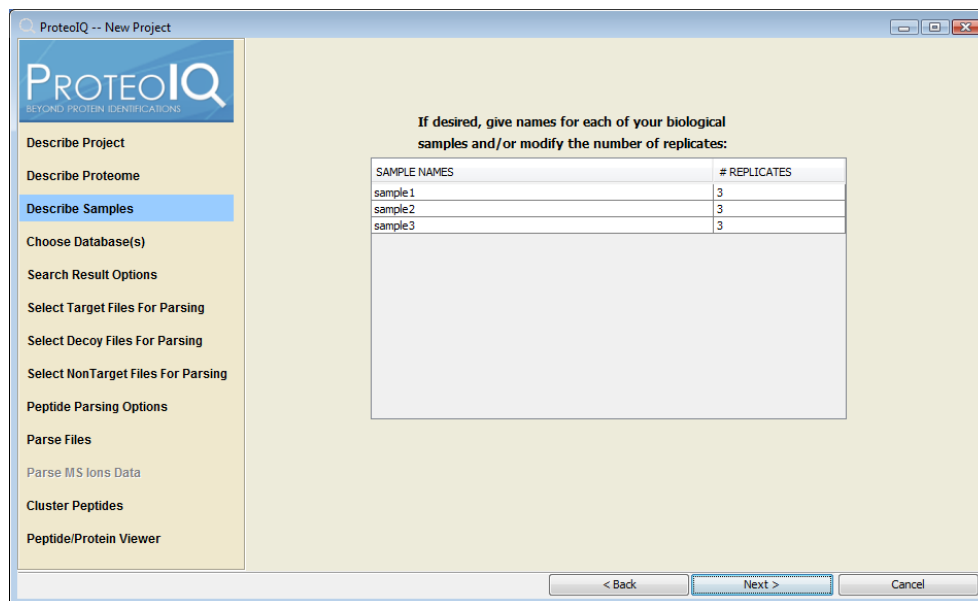
Each Result File Represents an Individual Sample

Performs protein identification in each database search results using **ONLY** peptides identified from that specific result file. The user should select this option, for example, when identifying spots from a 2-D gel and individual protein spots are excised, digested and analyzed independently. In this case, all of the peptides mapping to a protein would reasonably be expected to be in the same result file.

Describe Samples

Figure 27 shows the Describe Samples section of the New Project Wizard. Add **Sample Names** and modify the number of **Replicates**.

Figure 27. Described Samples



ProteoIQ -- New Project

PROTEOIQ
BEYOND PROTEIN IDENTIFICATION

Describe Project
Describe Proteome
Describe Samples
Choose Database(s)
Search Result Options
Select Target Files For Parsing
Select Decoy Files For Parsing
Select NonTarget Files For Parsing
Peptide Parsing Options
Parse Files
Parse MS Ions Data
Cluster Peptides
Peptide/Protein Viewer

If desired, give names for each of your biological samples and/or modify the number of replicates:

SAMPLE NAMES	# REPLICATES
sample1	3
sample2	3
sample3	3

< Back Next > Cancel

Sample Names

Inserting a unique name for each sample assists the user in identifying the individual sample sets in a complete proteome. This field is useful if biological replicates were analyzed, e.g., patient 1, 2, or 3. This field is optional as default sample names will be generated for the number of biological samples specified in the Describe Proteome page.

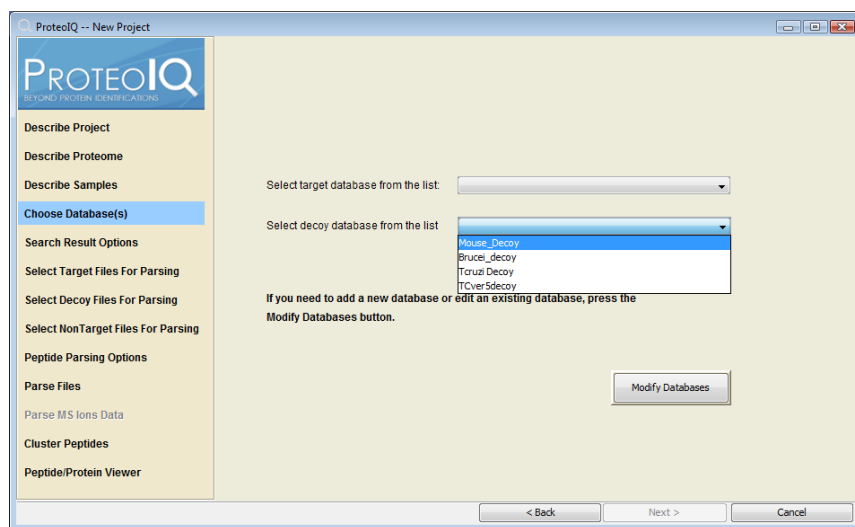
Replicates

Input the number of technical replicates performed on each biological replicate. This option is only available if Group File By Biological Sample is selected on the **Describe Proteome** page.

Choose Database

Select a target or decoy database using the drop down menus in the Choose Database(s) page (Figure 28). Database definitions may be modified or a new database may be created by selecting the Modify Databases button. To read more about creating databases in ProteoIQ see the Database Wizard tutorial.

Figure 28. Choose Database



Search Results Options

Select how peptides are used to assign proteins (Figure 29). Table 12 contains the options for selection in this section.

Figure 29. Search Results Options

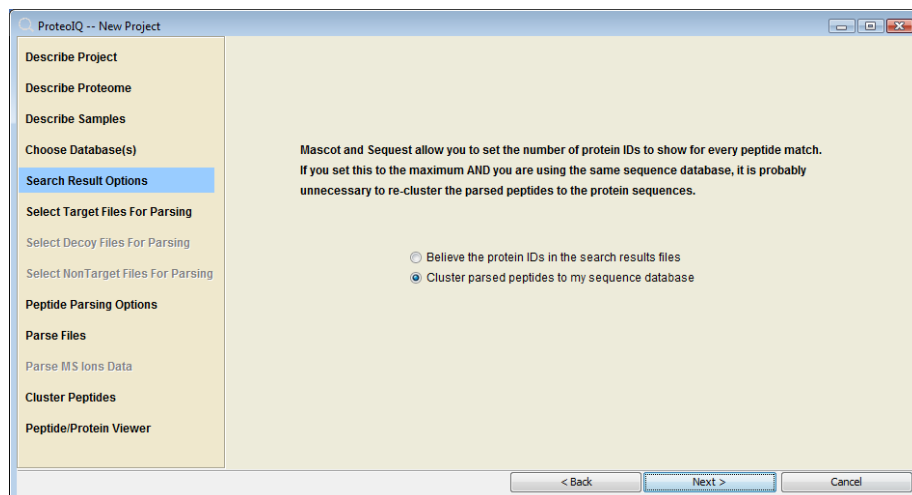


Table 12. Selections in Search Results Options

Field	Description
Believe the protein IDs in the search results files	Reports ONLY proteins assigned to peptides in the database search result. If this is selected peptides are not mapped to the FASTA database. For more information see “ Believe The Protein IDs in the Search Results Files ”.
Cluster parsed peptides to my protein database	By selecting this option, ProteoIQ will align all peptide matches to the protein sequences within the specified database. This is the recommended option for most proteome data sets.

Believe The Protein IDs In The Search Results Files

Reports ONLY proteins assigned to peptides in the database search result. If this is selected peptides are not mapped to the FASTA database.

Database search software allows the user to set the number of proteins identifications displayed for every peptide match. Since ProteoIQ arranges protein identifications into Protein Groups each peptide match is assigned to all the proteins in the specified database that the database search software determines contains this sequence.

If the maximum number of proteins displayed for each peptide match in the database search software is set too low, then protein group creation will not be accurate. Alternatively, if the user’s database does not contain a large number of homologous protein families and the number of proteins displayed for each peptide match in database search software is set to the maximum, then it may be unnecessary to perform the clustering.

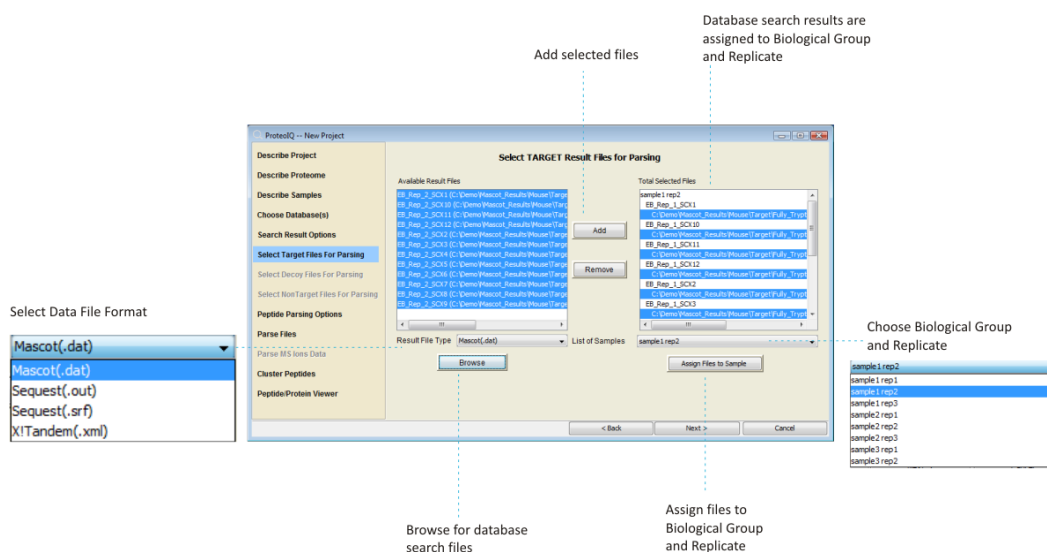
Warning: If clustering is not performed, then the parse rules defined in ProteoIQ for your database **MUST** correlate with the rules used in the database search software. Otherwise, the protein identifications will not be properly extracted from the database search results. For this reason we recommend **ALWAYS** clustering.

Select Files for Parsing

The Select Files for Parsing window is used to add database search results and assign them to Biological Groups and Replicates (Figure 30). Please see the “Grouping Database Search Results” tutorial for a more detailed description of this process.

Result File Type allows you to select the database search results format to be grouped.

Figure 30. Select Files for Parsing



Grouping database search results in ProteoIQ is performed as follows:

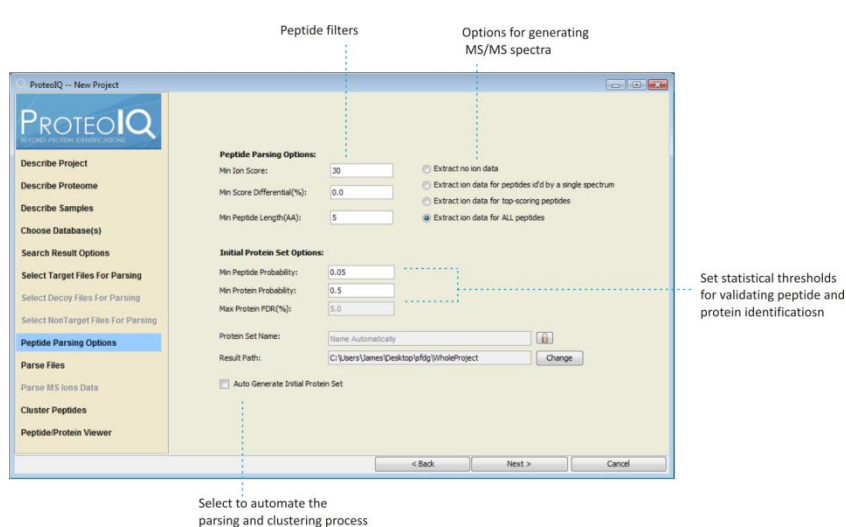
1. Select the search results format. Note: multiple search result formats can be imported into ProteoIQ within the same project.
2. Browse database search results. The name of the peak list used to produce the database search result is rendered in the Available Results Files viewer.
3. Select the files for grouping.
4. Once selected, the files appear as highlighted. Move the highlighted files to the Total Selected Files viewer by selecting Add.
5. Select the appropriate result file(s) in the Total Selected Files viewer
6. Select the Biological Group or Replicate from the List of Samples drop down menu.
7. Click Assign Files to Sample.

The assigned results files will then appear as tabbed under the Biological Group name in the Total Selected Files viewer. At any time the results files may be removed or ungrouped by highlighting either the result file or group name and selecting the remove button.

Manually group search result files to source – if selected, allows source files with different names to be grouped together as if from the same source (MS/MS file).

Peptide Parsing Options

The Peptide Parsing Options window as shown in **Figure 31**, is used to set the parameters for creation of the first protein set. Apply filters to the data and select parameters for creating **MS/MS spectral views** during peptide parsing. Table 13, Table 14, and Table 15 contain descriptions of each parameter.



31. Peptide Parsing Options

There are five components

to the Peptide Parsing Options:

1. **Peptide Parsing Options - Peptide filters**
2. **Initial Protein Set Options - Protein and peptide statistical thresholds**
3. **Options for generating MS/MS spectral views**
4. **Specifying name and location for ProteoIQ Project**
5. **Choose auto protein set creation or manual**

Table 13. Peptide Parsing Options

Field	Description
Min Score	Removes peptides below the Mascot Ion Score, SEQUEST XCorr, and X!Tandem Hyper Score.
Min Score Differential	Removes peptides below the % difference (in score) between the most significant match and the second best match to a given spectra.
Min Peptide Length	Removes peptides having fewer amino acids than the setting.

Table 14. Initial Protein Set Options

Field	Description
Min Peptide Probability	Removes peptides below the specified Peptide Probability.
Min Protein Probability	Removes proteins below the specified Protein Probability.
Min Protein FDR	Removes proteins above the specified Protein False Discovery Rate.

Table 15. Options for MS/MS Spectral Creation and other parameters

Field	Description
Create No Ion Data	This is the default setting for ProteoIQ and will NOT create any MS/MS spectra. This setting will create the smallest possible ProteoIQ project.
Extract Ion Data For Peptides ID'd From a Single Spectrum	Creates an MS/MS spectrum for a peptide identification that was generated from a single MS/MS event (i.e. with a spectral count of one).
Create Ion Data For Top Scoring	Creates MS/MS spectra for top scoring peptides ONLY . The most significant match (by score) for each peptide will be linked with an MS/MS spectrum.
Create Ion Data For All Peptides	Generates an MS/MS spectrum for every peptide identification regardless of score or rank.
Protein Set Name	Defines the name of the first protein set.
Results Path	Specifies the location of the ProteoIQ Project.
Auto Generate Initial Protein Set	Automates the process of Peptide Parsing and Cluster. If not selected the user will have to manually initiate Peptide Parsing and Clustering.

Min Score

The user may set a minimum score threshold for extraction of peptides from the database search results. For **Mascot** the score threshold refers to **Ion Score**. For **SEQUEST** score indicates **XCorr** and when **X!Tandem** results are used, score refers to **Hyper Score**.

Min Score Differential (%)

The Min Score Differential function allows the user to select the degree of difference between the most significant match and the second best match to a given spectra. For SEQUEST users this setting is loosely analogous to the delta Cn score. If the top match passes the user-defined threshold, then the assignment to that spectra will be included in the final peptide list. However, if the #1 match does not have a score (**Ion Score**, **Xcorr** or **Hyper Score**) that is higher than the #2 match by the percentage set in the Min Score Differential, then the assignment to that spectra will not be included in the final peptide list. See Example 1 for how Min Score Differential is applied in ProteoIQ.

Min Score Differential and Multiple DB Search Results for a Single Peak List

ProteoIQ supports an unlimited number of database search results for a single peak list. One common example is to perform two db searches first using fully tryptic enzyme specificity, and then performing a second search using partially tryptic or no enzyme specificity. All database search results may be loaded in simultaneously and ProteoIQ will determine the best match for each query (spectrum) between all db search results. If a Min Score Differential is applied, ProteoIQ will compare the Scores between each search result and determine if the #1 match passes the user defined Min Score Differential.

Example 1. Min Score Differential

Background

Query #399 was matched to two peptides P1 and P2.

P1: IAYNVEAAR (Ion Score of 58.4)

P2: TGGNKTEAAR (Ion Score of 18.4).

Set **Min Score Differential** to **20%**.

This means that the #1 match (IAYNVEAAR) must have a score that is 20% higher than the #2 match (TGGNKTEAAR) to be included in the peptide list following search results parsing.

ProteoIQ performs this calculation as follows:

T = the score of the #1 match

NT = the score of the #2 match

min Δ = % min score differential

$T \geq [NT + (T \times \text{min } \Delta)]$

So: $18.4 + (58.4 \times 0.2) = 30.08$

Since 58.4 is ≥ 30.08 then the peptide identification is included.

Another way to look at this: the #1 match (58.4) must have a score greater than 46.72 ($58.4 - (58.4 \times 0.2)$) to be 20% higher than the #2 match.

Min Peptide Length

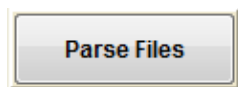
Enter the number of amino acids that must be present in an identified peptide for it to be included in the final peptide list. For example: if the Min Peptide Length is set to 5, then any peptides with 4 or less amino acids will be removed.

Warning: Low scoring peptide matches are required for accurate calculation of peptide probabilities by Peptide Prophet; therefore, if probability thresholds are to be applied, then the user should consider **NOT** applying filters during the parsing step.

Parse Files\MS Ions Data

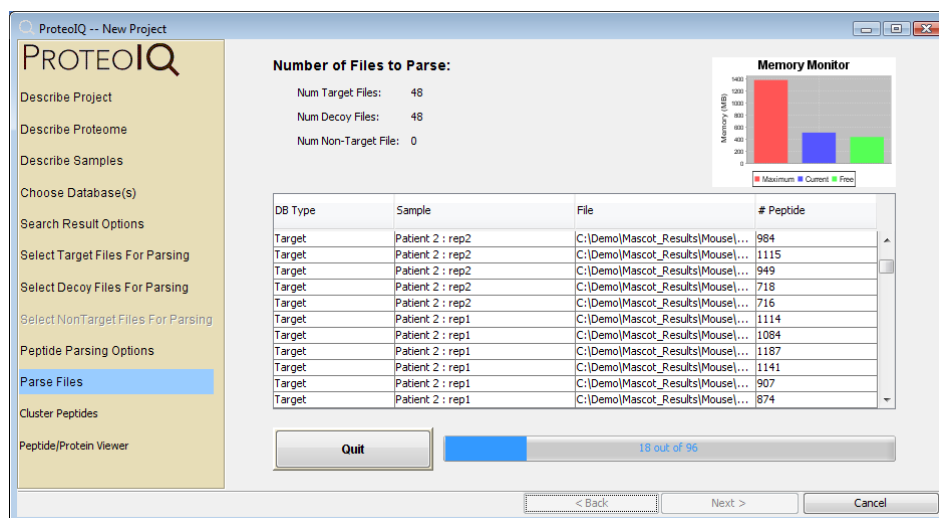
The parse files progress window displays the information about each database search result as ProteoIQ extracts the relevant peptide and protein information (Figure 32).

Select



to initiate parsing of the database search results.

Figure 32. Parse Files Progress

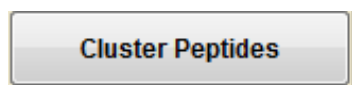


Note: For large data sets (especially LTQ results) the parsing process will take some time. In these cases, we recommend running the parsing and clustering steps in auto mode by selecting Auto Generate Initial Protein Set.

Cluster Peptides

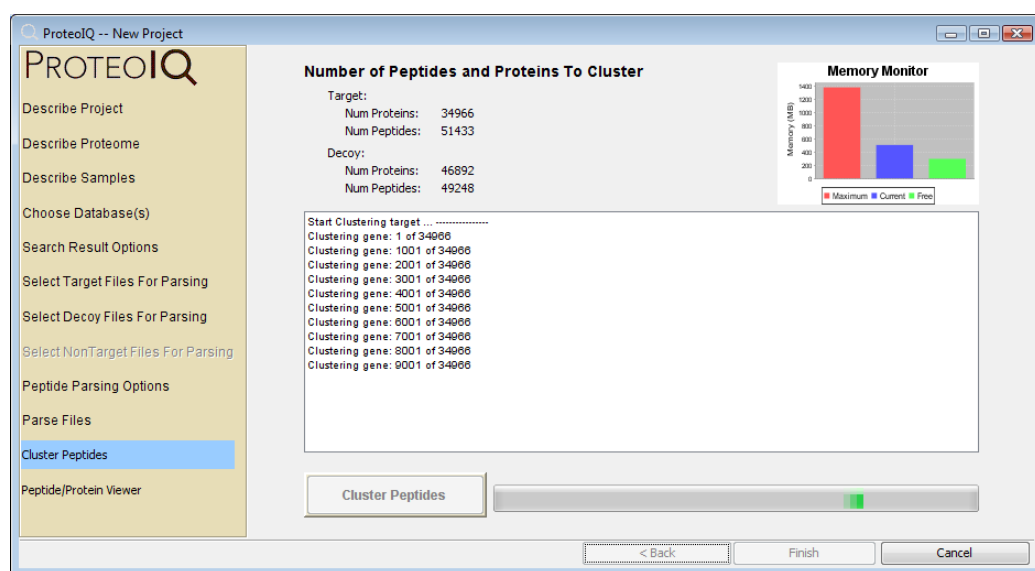
Figure 33 shows the cluster Peptides progress window displays the status of mapping peptide to a FASTA database during the creation of a new ProteoIQ Project.

Select



to initiate peptide to protein clustering.

Figure 33. Cluster Peptides Progress Window

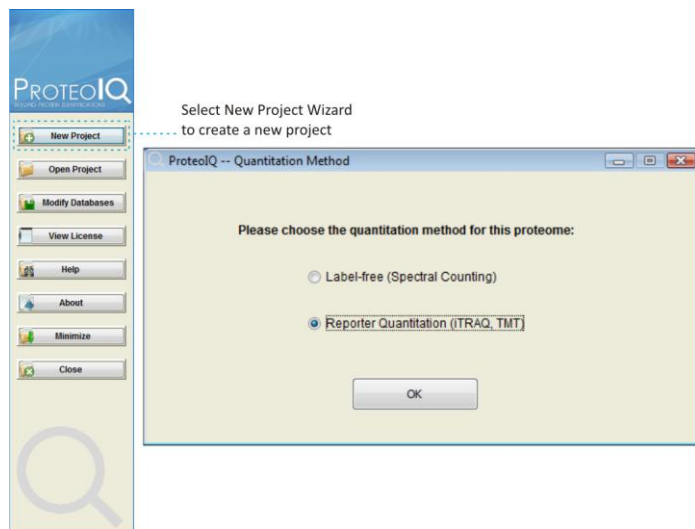


Note: For large databases the clustering process will take some time. In these cases, we recommend running the parsing and clustering steps in auto mode by selecting Auto Generate Initial Protein Set.

Creating a New Project – Reporter Ion Quantitation

To create a new project in ProteoIQ, first select the New Project Wizard from the Navigation Pane as shown in Figure 34, then select Reporter Ion Quantitation.

Figure 34. New Project Wizard



There are seventeen sections in the [New Project Wizard](#). The following describe the individual sections in the [New Project Wizard](#).

- 🔍 [Describe Project](#)
- 🔍 [Describe Proteome](#)
- 🔍 [Describe Experiment](#)
- 🔍 [Describe Correction Factor](#)
- 🔍 [Describe Samples](#)
- 🔍 [Map Ions to Groups](#)
- 🔍 [Choose Database](#)
- 🔍 [Search Results Options](#)
- 🔍 [Select Files for Parsing](#)
- 🔍 [Select Label Modification](#)
- 🔍 [Peptide Parsing Options](#)
- 🔍 [Parse Files/MS Ions Data](#)
- 🔍 [Cluster Peptides](#)
- 🔍 [Peptide/Protein Viewer](#)

Describe Experiment

The Describe Experiment section contains fields for specifying how the reporter quantitation experiment was performed. Select the reporter method, specify a custom method, set replicates and multiplicity. Figure 35 shows the Describe Experiment section of the [New Project Wizard](#). Table 16 describes the reporter methods and Table 17 describes the fields.

Figure 35. Describe Experiment Section

ProteoIQ -- New Project -- Reporter Quantitation

Please describe your experiment:

Choose Reporter Method:

☐ iTRAQ2 ☐ TMT2 ☐ Custom

☒ iTRAQ4 ☐ TMT6 Method Name:

☐ iTRAQ8 Number of Ions:

Experiment Multiplicity: Set multiplicity level

Number of Replicates: Specify number of replicates

Replication Method:

☒ External (1 group per label, replicates are separate results) Assign internal or external replicates

☐ Internal (1 replicate per label, 1 set of results)

Mass Tolerance(da) for matching reporter ions in each spectrum: Mass tolerance for labeling reporter ions

< Back Next > Cancel

Table 16. Reporter methods supported in ProteoIQ

Reporter Method	Reporter Masses Used for Quantitation
iTRAQ2 (Duplex)	114.1, 117.1
iTRAQ4 (Four plex)	114.1, 115.1, 116.1, 117.1
iTRAQ8 (Eight plex)	113.1, 114.1, 115.1, 116.1, 117.1, 118.1, 119.1, 121.1
Tandem Mass Tag (TMT2)	126.0, 127.0
Tandem Mass Tag (TMT6)	126.0, 127.0, 128.0, 129.0, 130.0, 131.0

Table 17. Fields in the Describe Experiment Section

Field	Description
Choose Report Method	Select the reporter ion used for your experiment (iTRAQ, TMT, or Custom) Note: The custom report ion selection allows the user to input data from any report ion (whether commercial or in-house) that was utilized in an experiment. Neither the mass nor the number of ions is limited.
Experimental Multiplicity	Multiplicity indicates that multiple reporter ion experiments were performed using a single label with a greater number of samples than the number of reporter ions utilized. Example: A researcher wants to analyze three cell types, but only has an iTRAQ2 reagent. Multiplicity allows him/her to do perform two MS experiments. Experiment one contains cell type one and cell type two and experiment two contains cell type one and cell type three. With a multiplicity of two, ProteoIQ will compare both cell type two and three to cell type one for quantitation. Note: The only restriction is that the control be the same for all experiments. Therefore, one can make comparison across multiple Plexes on an “apples to apples” basis.
Number of Replicates	Indicates the number of replicates performed. These can be either internal or external.
Replication Method	Select whether replication was done internally or externally. Internal replication example: iTRAQ 4 reagent is used to analyze two samples, in duplicate. External replication indicates that multiple MS/MS experiments were performed for the same set of samples.
Mass Tolerance (DA)	Indicates the mass window used for selection of the reporter ions. Note: A Mass Tolerance of 0.1 for a reporter ion mass of 114.1 indicates that ProteoIQ will search for the reporter ion from 114.0 to 114.2.

Describe Correction Factor

Correction factors are designed to adjust for batch-to-batch deviations in purity of the reporter reagents. Purity coefficients are typically supplied by the reagent manufacturer to indicate the percentages for each reporter ion that differ by -2, -1, +1, and +2 from the mass of each reporter ion. ProteoIQ will automatically adjust the peak intensities of each reporter ion based on the correction factors selected, prior to quantitation.

Figure 36 shows the Describe Correction Factor section of the [New Project Wizard](#).

Figure 36. Describe Correction Factor

Select previously saved correction factors

Please choose from existing reagent correction factor settings or create settings by unlocking the table and entering correction factors for your reagent below.

Stored correction factor settings: Demo Settings

Label Peak	% of -2	% of -1	% of 0	% of +1	% of +2
114.1	0.0	1.0	92.5	6.3	0.2
115.1	0.0	2.0	91.9	6.0	0.1
116.1	0.0	3.0	92.1	4.9	0.0
117.1	0.0	3.8	92.2	3.8	0.1

☐ Lock Settings Settings Name: Demo Settings

Delete Settings Save Settings

< Back Next > Cancel

Edit settings

Name new settings

Save settings

The following steps outline how to create a set of correction factors:

1. Deselect Lock Settings
2. Enter a name for the new correction factor in the Settings Name field
3. To edit a cell, double click on the cell and type in the desired value
4. Click Save Settings to save the correction factors

Describe Samples

Figure 37 shows the Describe Samples section of the [New Project Wizard](#). Add Samples Names and select the sample to be used as the reference group.

Figure 37. Described Samples

ProteoIQ -- New Project -- Reporter Quantitation

Describe Project
Describe Proteome
Describe Experiment
Describe Correction Factor
Describe Samples
Map Ions to Groups
Choose Database(s)
Search Result Options
Select Target Files For Parsing
Select Decoy Files For Parsing
Select NonTarget Files For Parsing
Select Label Modification
Peptide Parsing Options
Parse Files
Parse MS Ions Data
Cluster Peptides
Peptide/Protein Viewer

Please provide names for your biological groups and modify the number of replicates (if necessary):

SAMPLE NAMES	# REPLICATES	REFERENCE GROUP
Control	2	☒
Cell 1	2	☐
Cell 2	2	☐
Cell	2	☐

< Back Next > Cancel

Sample Names

Inserting a unique name for each sample assists the user in identifying the individual sample sets in a complete proteome. This field is useful if biological replicates were analyzed, e.g., patient 1, 2, or 3. This field is optional as default sample names will be generated based on the [Reporter Method](#) selected.

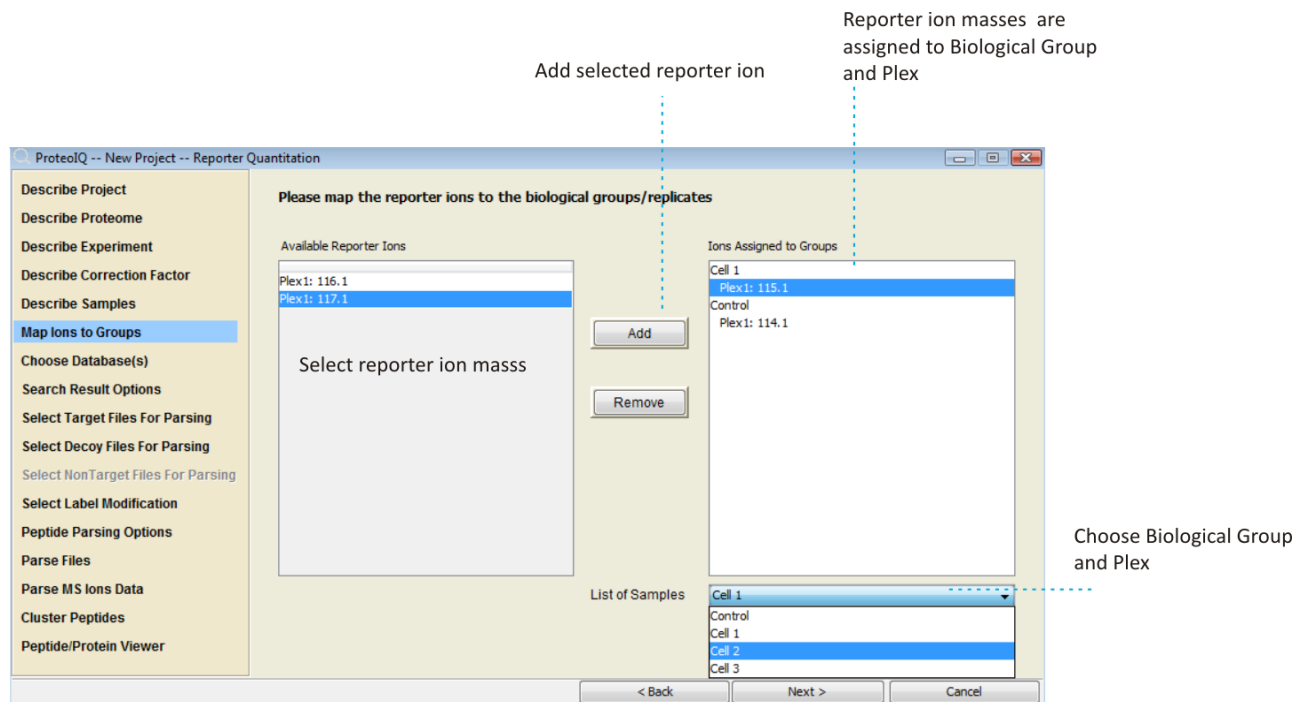
Reference Group

When a reference group is selected the reporter ion corresponding to the reference group will be used as the “control” and will act as the denominator in the ratio calculations. For example: if an iTRAQ duplex experiment is performed and the 114.1 is set as the reference group then quantitation will be performed by taking the ratio of reporter ion intensities as 117.1/114.1.

Map Ions to Groups

The Map Ions to Groups window is used to associate reporter ion masses with a **Biological Group** and **Plexes** (Figure 38).

Figure 38. Map Ions to Groups



Mapping reporter ions to biological groups in ProteoIQ is performed as follows:

1. Select the reporter ion mass and plex in the Available Reporter Ions viewer.
2. Once selected, the reporter ion masses appear as highlighted. Move the highlighted reporter ions to the Ions Assigned to Groups viewer by selecting Add.
3. Select the appropriate reporter ion mass in the Ions Assigned to Groups viewer
4. Select the Biological Group from the List of Samples drop down menu.
5. Click Assign Ion to Sample.

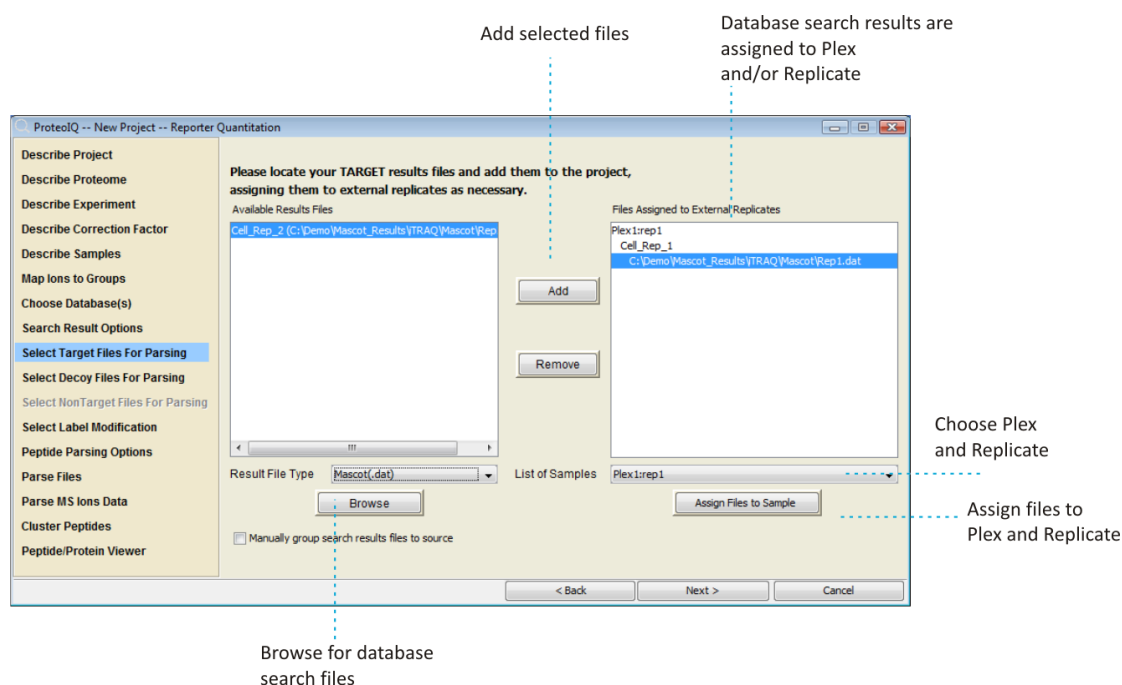
The assigned reporter ion masses will then appear as tabbed under the Biological Group name in the Ions Assigned to Groups viewer. At any time the reporter ions may be removed or ungrouped by highlighting either the reporter ion mass or group name and selecting the remove button.

Select Files for Parsing

The Select Files for Parsing window is used to add **Target**, **Decoy**, or **Non-Target** database search results and assign them to Replicates and Plex (Figure 39). Please see the “Grouping Database Search Results” tutorial for a more detailed description of this process.

Result FileType allows you to select the database search results format to be grouped.

Figure 39. Select Files for Parsing



Grouping database search results in ProteoIQ is performed as follows:

1. Select the search results format. Note: multiple search result formats can be imported into ProteoIQ within the same project.
2. Browse database search results. The name of the peak list used to produce the database search result is rendered in the Available Results Files viewer.
3. Select the files for grouping.
4. Once selected, the files appear as highlighted. Move the highlighted files to the Files Assigned to External Replicates viewer by selecting Add.
5. Select the appropriate result file(s) in the Files Assigned to External Replicates viewer
6. Select the Replicate and Plex from the List of Samples drop down menu.
7. Click Assign Files to Sample.

Manually group search result files to source – if selected, allows source files with different names to be grouped together as if from the same source (MS/MS file).

Select Label Modification

The Select Label Modification window allows the user to select which modification is associated with the reporter ion being used for quantitation (i.e. in the case of iTRAQ, the modification would be +144.1). To select the modification, simply highlight the correct modification mass and name then select Next.

Figure 40. Select label modification

ProteoIQ -- New Project -- Reporter Quantitation

Describe Project
Describe Proteome
Describe Experiment
Describe Correction Factor
Describe Samples
Map Ions to Groups
Choose Database(s)
Search Result Options
Select Target Files For Parsing
Select Decoy Files For Parsing
Select NonTarget Files For Parsing
Select Label Modification
Peptide Parsing Options
Parse Files
Parse MS Ions Data
Cluster Peptides
Peptide/Protein Viewer

Please select the modification associated with the labeling reagent in this experiment

Note: if possible, ProteoIQ auto-selects the most-likely modification

Modification Name	Modification Mass
+144.1	144.098068
+15.99	15.994904
+0.98	0.984009
enter the name of modification	enter the mass of modification

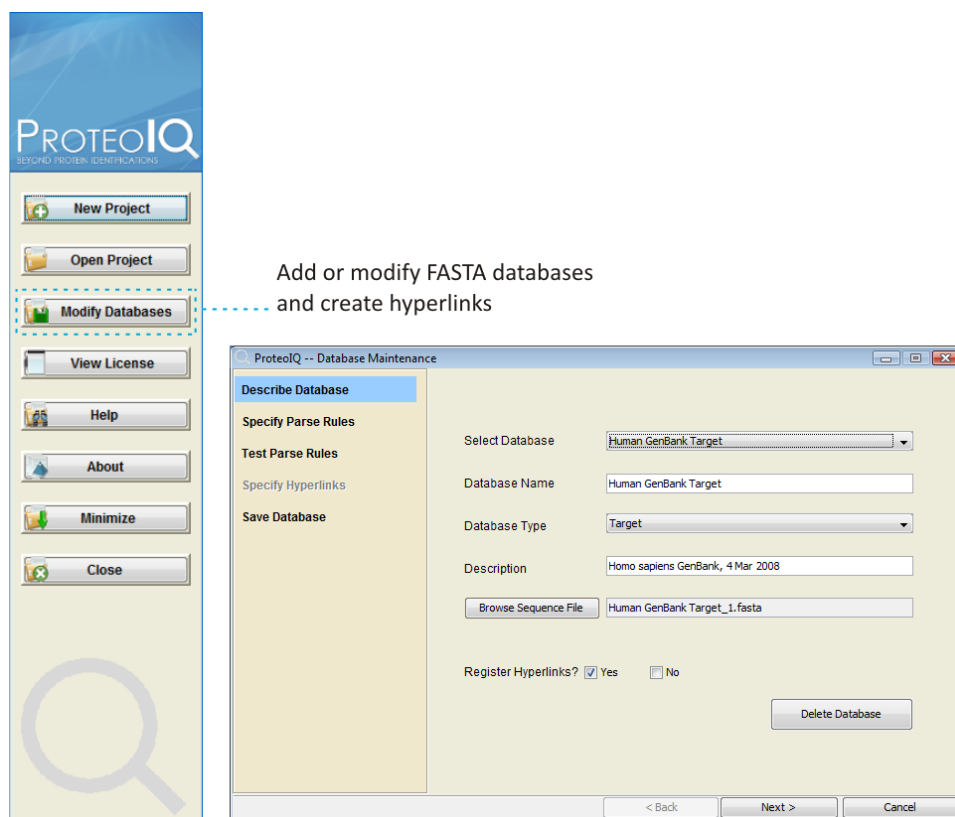
< Back Next > Cancel

Note: If the correct mass/name is not displayed, the user can enter the appropriate mass in the “enter the mass of modification” field.

Modify Database Wizard

To load or modify a **FASTA** database in ProteoIQ, first select the Modify Database Wizard from the Navigation Pane as shown in Figure 41. For a detailed description for creation of new databases please see the “Database Wizard” and “Creating New Hyperlinks” Tutorials.

Figure 41. Accessing the Modify Database Wizard



There are five sections in the Modify Database Wizard. The following describe the individual sections in the Modify Database Wizard.

- 🔍 **Describe Database**
- 🔍 **Specify Parse Rules**
- 🔍 **Test Parse Rules**
- 🔍 **Specify Hyperlinks**
- 🔍 **Save Database**

Describe Database

The Describe Database section contains fields for adding information about the new database. The information is stored when the database is saved and can be accessed through the Database Wizard at anytime by selecting the database name in the drop down menu. Figure 42 shows the Describe Database section of the Modify Database Wizard. Table 18 describes the fields.

Figure 42. Describe Database

The screenshot shows the 'ProteoIQ -- Database Maintenance' window. On the left is a sidebar with a tree view containing 'Describe Database' (selected), 'Specify Parse Rules', 'Test Parse Rules', 'Specify Hyperlinks', and 'Save Database'. The main area is titled 'Describe Database' and contains the following fields:

- Select Database:** A dropdown menu with 'Human GenBank Target' selected.
- Database Name:** A text field containing 'Human GenBank Target'.
- Database Type:** A dropdown menu with 'Target' selected.
- Description:** A text field containing 'Homo sapiens GenBank, 4 Mar 2008'.
- Browse Sequence File:** A button next to a text field containing 'Human GenBank Target_1.fasta'.
- Register Hyperlinks?:** A checkbox labeled 'Yes' is checked, and a 'No' checkbox is unchecked.
- Delete Database:** A button at the bottom right.

At the bottom of the window are three buttons: '< Back', 'Next >', and 'Cancel'.

Table 18. Describe Proteome

Field	Description
Select Database	The drop down menu allows the selection of a previously indexed database or the creation of a new database.
Database Name	Enter the name of the new database in the text field provided. Once indexed this name will be available in the select database drop down menu.
Database Type	Specify if the database is a separate target or decoy database or a concatenated target/decoy database. Decoy databases are used to calculate False Discovery Rates.
Description	A description of the database can be provided in the description text box. Text entered here will be stored with the indexed database.

Specify Parse Rules

The Specify Parse Rules section allows regular expressions to be created and saved to instruct ProteoIQ as to how to extract the accession numbers, protein names and descriptions from the FASTA database (Figure 43). Table 19 describes the fields.

Figure 43. Specify Parse Rules

ProteoIQ -- Database Maintenance

Describe Database
Specify Parse Rules
Test Parse Rules
Specify Hyperlinks
Save Database

Please specify the parse rules to extract the protein IDs and names from the sequence database.

Select Existing: 1 : DEFAULT

Description: DEFAULT

Protein ID Parse Rule: ^([^\s]+)

Protein Name Parse Rule: ^([^\s]+\s|.*)

Decoy Identifier Rule: REV|

< Back Next > Cancel

ProteoIQ comes preloaded with parse rules for commonly used database structures and custom parse rules can be created as follows:

Steps for creating custom parse rules:

1. Assign a name to the new parse rule in the Description field
2. Enter Protein ID parse rule
3. Enter Protein Name parse rule
4. Select Next to test the rule
5. Save database to permanently store the new parse rule

Table 19. Specify Parse Rules

Field	Description
Select Existing	Existing database parse rules can be accessed via the drop down menu or a new set of parse rules can be created.
Description	A description of the parse rule can be provided in the description text box. Text entered here will be stored with the parse rule once saved.
Protein ID Parse Rule	The regular expression that defines how to parse an accession string or protein ID from the fasta header line.
Protein Name Parse Rule	The regular expression that defines how to parse the protein name and description string from the fasta header line.
Decoy Identifier Rule	For use with concatenated databases ONLY. When creating a concatenated target and decoy database the user must specify a unique identifier to the decoy sequences. One example would be to add the term “Rev” to all protein names for the decoy sequences in the FASTA db. If “Rev” is entered in the Decoy Identifier Rule field then all peptides matching proteins with the name “Rev” will be treated as decoy matches for the FDR calculations. For more information on creating concatenated target/decoy databases see the Database Wizard Tutorial.

Test Parse Rules

The Test Parse Rules section displays the results of the [Parse Rule Settings](#). ProteoIQ applies the parse rules to the specified database and displays the Protein ID and Description fields. To ensure proper indexing of the database ensure that unique Protein IDs are being displayed for each protein entry. For [Concatenated Target + Decoy Databases](#), the Protein Type is displayed that indicates if the sequence is being considered a target or decoy based on the [Decoy Identifier Parse Rule](#).

Figure 44. Specify Parse Rules

Indicates if the sequence is a target or decoy as specified by the decoy identifier parse rule

ProteoIQ -- Database Maintenance

Describe Database

Specify Parse Rules

Test Parse Rules

Specify Hyperlinks

Save Database

Sequence file test results:

Database Name: Combined_T_G

FASTA Format Test: PASS

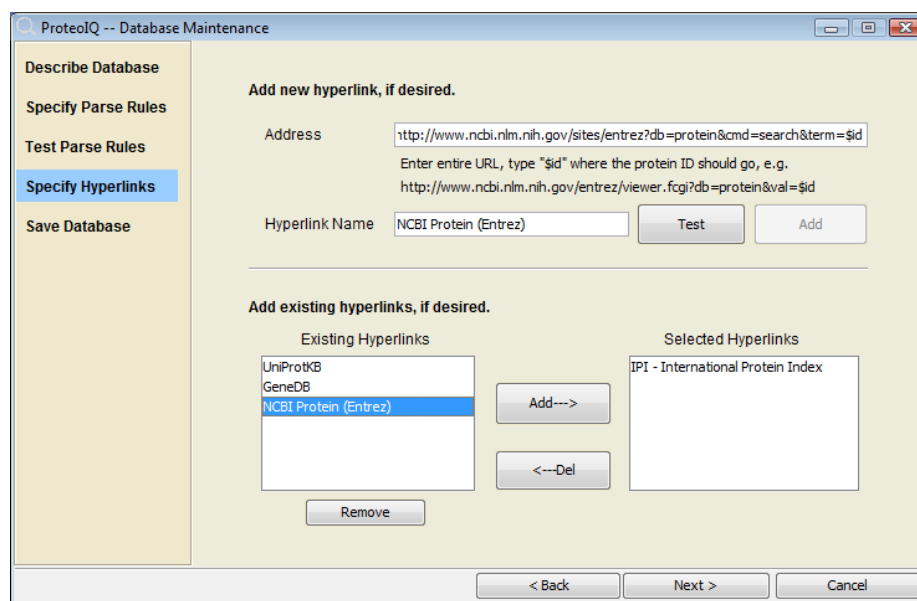
Protein ID	Description	Protein Type
parse rule: ^([^\s]+)	parse rule: ^([^\s]+)\s(.*)	gene type: TARGET/DECOY
CON_Trypsin	SWISS-PROT:P00761 TRYP_PIG Tr...	TARGET
CON_IP100174775.2	TREMBL:Q32MB2;Q86Y46 Tax_Id...	TARGET
CON_IP100290078.5	SWISS-PROT:P19013 Tax_Id=960...	TARGET
CON_00166205.2	TREMBL:Q7RTT2 Tax_Id=9606 Ge...	TARGET
CON_LysC	SWISS-PROT:P15636 Protease I p...	TARGET
REV_1703	Random Protein 1703	DECOY
REV_1704	Random Protein 1704	DECOY
REV_1705	Random Protein 1705	DECOY
REV_1706	Random Protein 1706	DECOY
REV_1707	Random Protein 1707	DECOY

< Back Next > Cancel

Specify Hyperlinks

The Specify Hyperlinks window allows hyperlinks to be added to a database during indexing. Once a protein set is created, the hyperlinks will appear as links from the protein name the registered database specified by the hyperlink. Figure 45 shows the Specify Hyperlink window. For a detailed description of how to create new hyperlinks, please see the “Create New Hyperlinks” tutorial.

Figure 45. Specify Hyperlinks



Steps to add an existing hyperlink:

1. Select hyperlink name in the Existing Hyperlinks field
2. Select Add
3. Click Next

Note: For a detailed tutorial on creating new hyperlinks in ProteoIQ please read the **Create New Hyperlinks** tutorial. Tutorials are located in the C:\Program files\ProteoIQ\docs path.

Using the Filter Pane

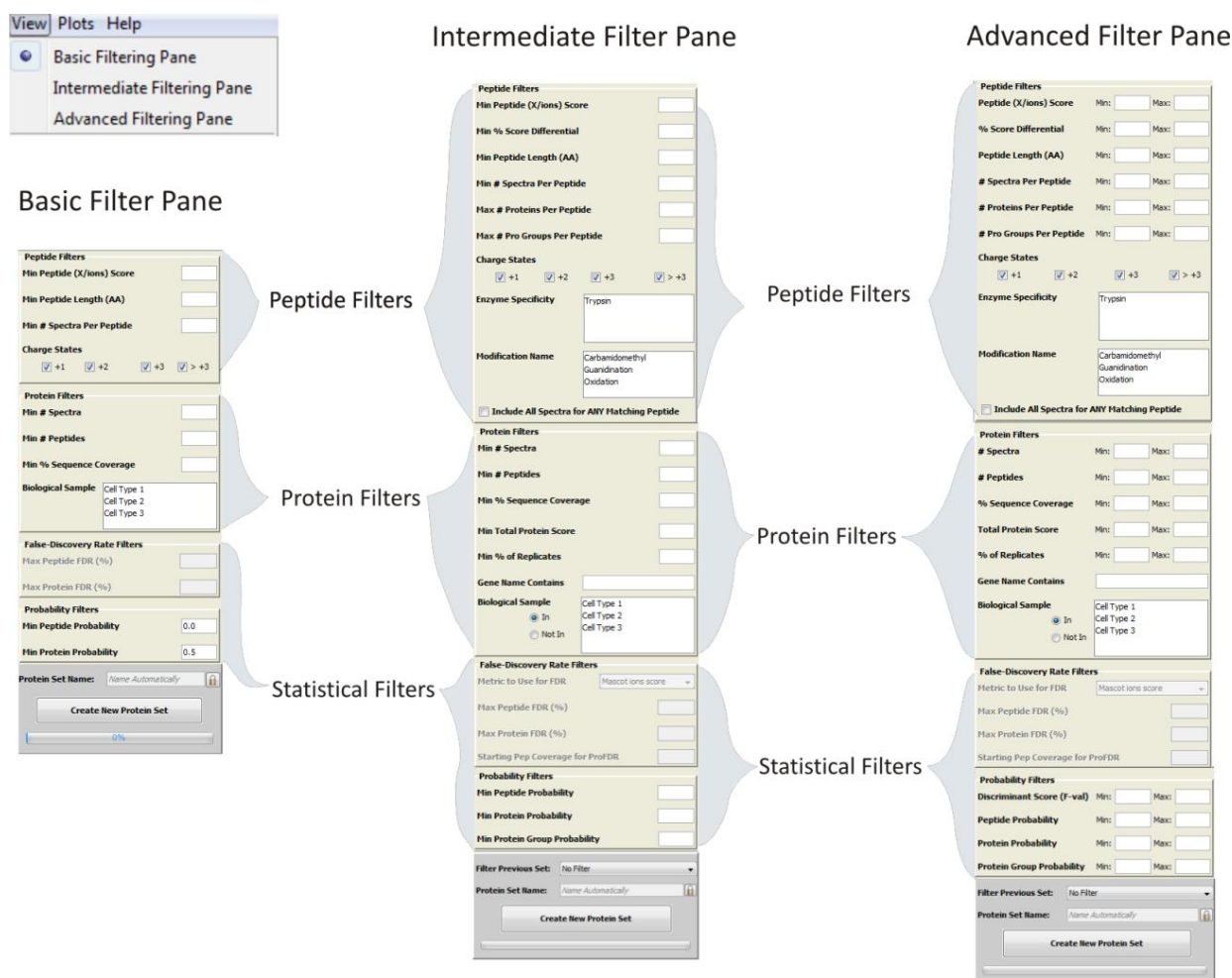
The filter pane is designed to allow the user to select a set of parameters for filtering a peptide and protein set. Topics in the section are:

- 🔍 [Selecting Filter Panes](#)
- 🔍 [How Filtering Works](#)
- 🔍 [Selecting Protein Sets to Filter](#)
- 🔍 [Description of Filter Parameters](#)

Selecting Filter Panes

There are three filter panes to choose from, as shown in Figure 46. To select a Filter Pane choose: **View > Basic, Intermediate, or Advance Filtering Pane**

Figure 46. ProteoIQ Filter Panes



How Filtering Works

The filtering pane is broken down into three sections.

1. Peptide Filters
2. Protein Filters
3. Statistical Filter

When a filter is applied the peptide and protein sets are filtered according to the parameters set in the Filter Pane and new peptide or protein set is displayed and can be accessed via the [Protein Set Navigation Bar](#).

Filters are applied in the following order:

1. If a [Peptide False Discovery Rate](#) is assigned, peptides are removed based upon the minimum [Xcorr](#), [F value](#), [Ion Score](#), or [Hyper Score](#) required to meet the user specified false discovery rate.
2. Peptides are then removed based upon the assigned peptide filters. For example, if a minimum peptide ion score of 40 is selected then all peptides below a score of 40 are first removed prior to protein assembly. If a peptide probability filter is assigned peptides are removed from the dataset based on the filter parameters.
3. Once [Protein Groups](#) are assembled, protein groups are selected based upon the user specified [Protein False Discovery Rate](#).
4. Protein filters are then applied to the proteins passing the [False Discovery Rate](#). Protein probability filters are applied at the same time as other protein filters such as #peptides or %sequence coverage.
5. If the maximum number of proteins groups matching a peptide is assigned then this filter is applied after the protein filtering.

Note: Filters are based off of the TOTAL RESULTS. Selecting a Biological Sample does NOT restrict the filters to the selected sample. In order to filter specific Biological Samples you must use the Copy Selected to New Protein Set Feature.

Selecting Protein Sets to Filter

Filters can be applied to an existing protein set or to all of the peptides extracted from the database search results (No Filter) as shown in Figure 47. Table 20 describes the fields in the Selecting Protein Sets Filter window.

Figure 47. Protein Set Selection Window for Filtering

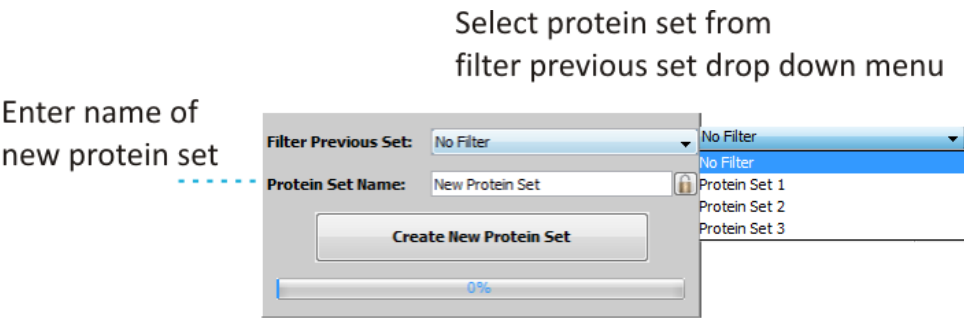


Table 20. Fields in the Protein Set Section Window

Field	Description
Filtering an Existing Protein Set	To filter an existing protein set, select the protein set name from the Filter Previous Set drop down menu.
No Filter	Selecting No Filter means that the filter will be applied to ALL peptides parsed from your database search results. This allows peptides not present in the original protein set to be displayed by losing the filter conditions.
Protein Set Name	Text field for specifying the name of the new protein set. This name will appear in the Protein Set Navigation Bar.

Description of Filter Parameters

Field	Description
Peptide Score	Score refers to either Mascot Ion Score, SEQUEST Xcorr, or X!Tandem Hyper Score.
Peptide Length (AA)	Number of amino acids contained in the peptide.
# Spectra per peptide	Number of MS/MS sequencing events for unique peptide identification. Also known as spectral counts.
# Proteins per Peptide	Specifies the number of proteins a peptide can match to. When performing spectral count quantification peptides that match to more than one protein can affect the accuracy of the quantification. Setting the max number of proteins per peptide to 1 removes all peptides that appear in more than one protein.
# Protein Groups per Peptide	Specifies the number of protein groups a peptide can match to. For proteomes with many homologous proteins a single peptide can match to many protein groups. This means that the peptide is not unique to a single protein group. If the user restricts the number of protein groups per peptide to 1 they in effect display only “unique” peptides. In other words only peptides that are uniquely assigned to a single protein group are displayed.
Charge state	If the filter by charge state is selected, ProteoIQ removes all peptides that were not identified with the specified charge states. Note: Filtering by charge state only requires that at least one occurrence of the peptide identification is a result of the assigned charge state. All other peptide identifications will also be displayed as long as one of them has the specified charge state.
Enzyme Specificity	Selecting enzyme specificity limits the displayed peptides to those identified in the database search results from which the given enzyme cleavage rules were applied.

Field	Description
Modification Name	Displays peptides identified with the user selected modification.
DB Search Engine	Displays peptides identified from the selected DB search engine.
Total Protein Score	Filters proteins based on total score.
# Peptides	The number of non redundant peptides matching a given protein.
% Sequence Coverage	Calculated as the sum of the number of amino acids in all identified peptides divided by the total number of amino acids in the protein multiplied by one hundred.
# Spectra	Number of MS/MS sequencing events for all peptides assigned to a given protein. Also known as spectral counts.
% of Replicates	Filters protein identifications based on their identification in a percentage or replicates. For example, if 4 replicates were performed and a value of 75% or replicates was applied then the protein would have to be identified in 3 out of 4 replicates to be included in the final protein set.
Gene Name Contains	This function allows the user to assign a text search of the entire protein dataset. Proteins with names that contain the user specified text string will be rendered as a new protein set.
Biological Sample	Filter produces a protein set containing proteins identified in the selected biological samples. Two options are available. If "In" is selected the protein set will contain proteins identified in the selected sample. If "Not In" is selected the created protein set will contain proteins identified in the biological sample that was not selected.

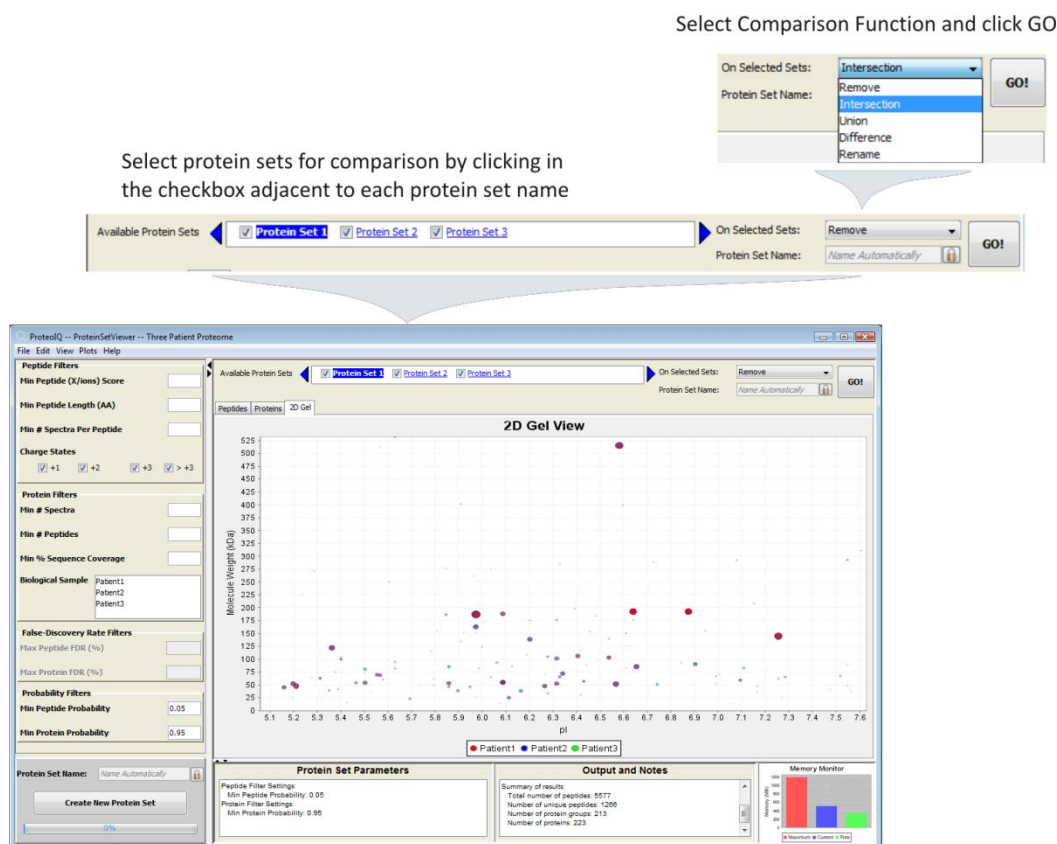
Field	Description
Metric to use for FDR	The drop down menu allows the user to select the type of peptide scoring function to utilize when ProteoIQ calculates the protein false discovery rate. Three choices are available: 1. Discriminate score (Fval) 2. Score for Mascot Ion Score, SEQUEST Xcorr, and X!Tandem Hyper Score. 3. Peptide probability
Max Peptide FDR (%)	Peptide false discovery rates are calculated by comparing the frequency of peptide identifications at each score between a target and decoy (reversed or scrambled) database. The false discovery rate is defined as the number of decoy peptides divided by the number of target peptides multiplied by 100%. When a max peptide FDR is defined, ProteoIQ selects peptide score thresholds such that the proportion of decoy peptide matches relative to the number of target peptide matches does not exceed the user defined max peptide FDR.
Max Protein FDR (%)	ProteoIQ utilizes the ProValT algorithm to calculate protein false discovery rates as described in Weatherly <i>et al.</i> ProValT assigns a protein false discovery rate based on the distribution of proteins with a minimum peptide coverage (c) identified by peptides score bins (B_s) between a target and decoy database search. By setting a max protein FDR (%) ProteoIQ selects peptide coverage levels and peptide score thresholds such that the proportion of decoy protein matches relative to the number of target protein matches does not exceed the user defined max protein FDR. Weatherly, DB., Atwood, JA., Minning, TA., Cavola, C., Tarleton, RL., Orlando, R. A heuristic method for assigning a false-discovery rate for protein identifications from Mascot database search results. <i>Mol. Cell. Proteomics</i>. 2005, 4, 762-772.

Field	Description
Starting Peptide Coverage	In ProteoIQ, when a protein false discovery is to be applied, the user must determine the starting peptide coverage (C) to use in the calculation of false discovery rates. The concept of a false discovery rate is based on the assumption that the null hypotheses (proteins identified in the random database) follow a quasi-normal distribution of minimal scores (S). Indeed, score distributions at increasing c may not necessarily look normal themselves, especially when the sample size is small. Thus, as the sample size (number of protein identifications) increases, for each c the score distribution will approach normality and the false discovery rate calculations will become more accurate. In practice determining C prior to protein false discovery rate analysis will be dependent on the distribution of minimal scores S in the target and decoy protein assignments. By default ProteoIQ begins the protein false discovery rate calculation using starting peptide coverage of 3. For large data sets the starting peptide coverage can be increased. Note: choosing correct starting peptide coverage is largely dependent on the frequency of protein identifications in the decoy database at that coverage level. After the protein groups have been rendered the user can evaluate the starting peptide coverage level by viewing the protein false discovery plot.
Discriminant score (F-value)	Filters peptides based on discriminant score or F-value.
Peptide Probability	Filters peptides based on peptide probability.
Protein Probability	Filters proteins based on total protein probability. Protein Probability takes into account the presence of protein groups and penalizes proteins if they contain peptides identified in multiple proteins within a group.
Protein Group Probability	Filters proteins based on total protein group probability. Protein Group Probability takes does not penalize proteins if they contain peptides identified in multiple proteins within a group.

Comparing Protein Sets

Comparing two or more protein sets in ProteoIQ is supported via the On Selected Sets drop down menu. Figure 48 shows that the On Selected Sets menu is located in the top right of the Peptide/Protein Set Viewer. Table 21 describes the functions.

Figure 48. Comparison Functions



Comparison functions are performed as follows:

1. Selecting protein sets to compare in the **Protein Set Navigation Bar** by checking the box adjacent to the Protein Set Name.
2. Select **Comparison Function** for On Selected Sets drop down menu.
3. Assign a name to the new protein set in the Protein Set Name text field.
4. Select GO to perform the function.

Table 21. Comparison Functions in the On Select Sets drop down menu

Field	Description
Remove	Deletes the selected protein set, including proteins and peptides.
Rename	To rename a protein set, select the protein set for renaming, enter text in the Protein Set Name field and select "GO!".
Intersection	Two or more protein sets must be selected. The intersection function renders a new protein set containing the proteins found in common between all of the selected protein sets.
Union	Two or more protein sets must be selected. The union function renders a new protein set that is the combined list of proteins from each of the selected protein sets.
Difference	Two or more protein sets must be selected. The difference function renders two or more protein sets that each contain the unique proteins found in each selected protein set. For example, if two protein sets are selected (protein set 1 and protein set 2) and the difference function is applied to both, two additional proteins sets will be created (protein set 3 and protein set 4). Protein set 3 will have proteins only found in protein set 1 and protein set 4 will have proteins only found in protein set 2.

Customizing Results Views

The peptide and protein tables in the Peptide and Protein Set Viewer can be modified to adjust the view. The following topics discuss how to customize the results views.

- 🔍 [Moving and Sorting Columns](#)
- 🔍 [Peptide Table Options in the Peptide Set Viewer](#)
- 🔍 [Protein Table Options in the Protein Set Viewer](#)
- 🔍 [Customize Protein Table in Protein Sequence Viewer](#)
- 🔍 [Customize Peptide Table in Protein Sequence Viewer](#)
- 🔍 [Modifying Plot Views](#)
- 🔍 [Selecting Proteins in View](#)

Moving and Sorting Columns

Columns in both the Peptide and Protein Set Viewer can be modified using three options (Figure 49).

1. Rearranged columns by dragging the column header.
2. Sort results by clicking on the column header.
3. Select **View > Peptide Table Options** or **View > Protein Table Options** to set which columns and rows are displayed.

Figure 49. Protein Set Viewer Customization

Select **View > Protein Table Options**

Select proteins tab to display identified proteins

All columns can be sorted or moved

Drag columns to change view

Protein Group	Is Top?	Sequence Id	Sequence Name	Total Score	Total Spectra	Patient1 Score	Patient1 Spectral Count	Patient1 N-SC Log2 Relative Expression
1	Yes	NP_000375.2	apolipoprotein B precursor [Homo sapiens]	434.631	496	253.063	129	-0.552
2	Yes	NP_000055.2	complement component 3 precursor [Homo sapiens]	411.292	836	265.714	230	-0.478
2	No	XP_001719515.1	PREDICTED: hypothetical protein, partial [Homo sapiens]	297.244	587	191.374	164	-0.458
2	No	XP_001724196.1	PREDICTED: similar to complement component 3 [Homo sapiens]	106.297	227	66.589	59	-0.558
3	Yes	NP_001002029.2	complement component 4B preproprotein [Homo sapiens]	210.802	373	165.427	108	-0.386
3	No	NP_009224.2	complement component 4A preproprotein [Homo sapiens]	210.802	373	165.427	108	-0.386
4	Yes	NP_000583.2	complement component 4B preproprotein [Homo sapiens]	204.916	353	157.907	104	-0.358
4	No	NP_001002029.1	complement component 4B preproprotein [Homo sapiens]	204.916	353	157.907	104	-0.358
5	Yes	NP_000087.1	ceruloplasmin precursor [Homo sapiens]	152.203	179	98.085	52	-0.375
6	Yes	NP_001701.2	complement factor B preproprotein [Homo sapiens]	104.856	150	49.674	44	-0.371
7	Yes	NP_000005.2	alpha-2-macroglobulin precursor [Homo sapiens]	99.673	128	41.95	24	-1.046
8	Yes	NP_001726.2	complement component 5 preproprotein [Homo sapiens]	96.593	104	52.028	32	-0.353
9	Yes	NP_000574.2	vitamin D-binding protein precursor [Homo sapiens]	88.866	120	45.9	28	-0.737
10	Yes	NP_000177.2	complement factor H isoform a precursor [Homo sapiens]	85.571	99	45.055	15	-1.342
10	No	NP_001014975.1	complement factor H isoform b precursor [Homo sapiens]	22.648	26	11.086	3	-1.038
11	Yes	NP_000604.1	hemopexin [Homo sapiens]	82.691	195	61.647	53	-0.481
12	Yes	NP_596869.3	titin isoform N2-A [Homo sapiens]	79.406	92	27.172	24	-0.604
13	Yes	NP_002209.2	inter-alpha (globulin) inhibitor H4 [Homo sapiens]	73.742	84	54.566	29	-0.141
14	Yes	NP_002207.2	inter-alpha globulin inhibitor H2 polypeptide [Homo sapiens]	65.591	86	37.901	26	-0.443
15	Yes	NP_001076.2	serpin peptidase inhibitor, clade A, member 3 precursor [Homo sapiens]	65.531	176	48.367	60	-0.201
16	Yes	NP_000884.1	kininogen 1 isoform 2 [Homo sapiens]	62.489	104	34.879	25	-0.677
16	No	NP_001095886.1	kininogen 1 isoform 1 [Homo sapiens]	55.964	87	28.353	17	-1.038
17	Yes	NP_570602.2	alpha 1B-glycoprotein precursor [Homo sapiens]	60.067	77	28.532	12	-1.386
18	Yes	NP_000497.1	coagulation factor II precursor [Homo sapiens]	59.524	60	28.717	15	-0.656
19	Yes	NP_000473.2	apolipoprotein A-I-V precursor [Homo sapiens]	58.287	81	32.426	16	-0.964
20	Yes	NP_002206.2	inter-alpha (globulin) inhibitor H1 [Homo sapiens]	57.771	69	31.107	11	-1.429

Peptide Table Options in Peptide Set Viewer

Figure 50 shows the Peptide Column selection menu that is accessible by selecting **View > Peptide Table Options > Select Columns**. Select columns to display by checking the box adjacent to each column name. Table 22 provides descriptions of each data type.

Figure 50. Peptide Column Selection Menu

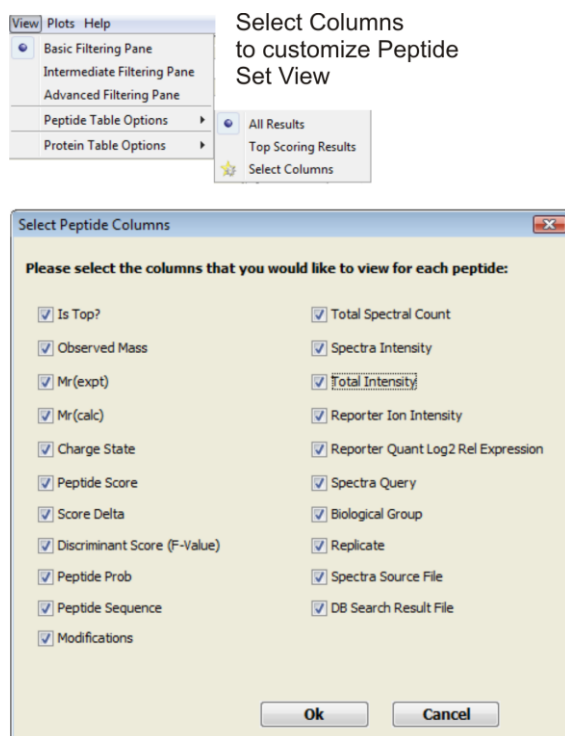


Table 22. Data Types Available in the Peptide Column Selection Menu

Field	Description
Is Top?	Indicates if a peptide assignment has the top Peptide Score. The same peptide is often identified multiple times. To display only the Top Scoring Peptides and remove redundancy, select Top Scoring Results in the Peptide Table Options.
Observed Mass	The observed peptide m/z that was selected to generate the MS/MS spectrum. Otherwise known as precursor m/z.
Mr(expt)	Calculated molecular mass from the observed mass. For example: Observed Mass = 1000.00 and Charge State = 3 then the Mr(expt) = 3000.00.

Field	Description
Mr(calc)	Theoretical molecular mass calculated from the sum of masses of the amino acids for each peptide.
Charge State	Charge states are not calculated in ProteoIQ. The charge states are extracted from the database search results.
Peptide Score	Mascot Ion Score, SEQUEST Xcorr, or X!Tandem Hyper Score.
Score Delta	Score difference between the 1 st and 2 nd best peptide assignments to a given MS/MS spectrum. The score delta is analogous to the Delta Cn in SEQUEST.
Discriminant Score (F-value)	Score normalization performed prior to probability calculation. For a detailed description of how F-values are calculated, see the Discriminant Score section.
Peptide Probability	The probability that the peptide assignment to a given spectra is correct.
Peptide Sequence	The amino acid sequence of the identified peptide.
Modifications	Peptides modifications are displayed in a separate column and are reported by name and amino acid position.
Total Spectral Count	If a peptide is identified multiple times the Total Spectral Count will equal the number of times the peptide was identified.
Spectra Intensity	For Mascot results this value is the intensity of the precursor ion that generated the MS/MS spectrum. In SEQUEST results, this value is the total ion current (Total Inten) of the MS/MS scan.
Total Intensity	The sum of the Spectra Intensities for all identical peptides.
Reporter Ion Intensity	Displays the intensity of each reporter ion. Note: for the intensity to be displayed the peptide MUST contain the appropriate modification and the reporter ion MUST be within the specified mass window.

Field	Description
Reporter Quant Log ₂ Rel Expression	Displays the Log ₂ expression for each peptide as calculated based on the intensity of the reporter ions.
Spectra Query	For Mascot results Spectra Query is the query number. For SEQUEST results Spectra Query equals the number of the .OUT file. In X!Tandem results the Spectra Query equals the Domain id.
Biological Group	Specifies the Biological Group the peptide identification resulted from.
Replicate	Specifies the Replicate the peptide identification resulted from.
Spectra Source File	Indicates the database search result files that the identified peptide originated from. (.pkl, .mgf, .DTA, .xml, etc.) that was used to perform the database search which lead to the peptide identification.
DB Search Result	Indicates the database search result files that the peptide identification originated from.

Note: When a ProteoIQ project is saved the Peptide and Protein Views will also be saved.

Protein Table Options in Protein Set Viewer

Figure 51 shows the Protein Column selection menu that is accessible by selecting **View > Protein Table Options > Select Columns**. The Select Protein Columns window will be dependent on the type of quantification performed; either Label Free or Reporter Ion Quantitation. Select the columns to display by checking the box adjacent to each column name. Table 23Table 24, Table 25, and Table 26 provides descriptions of each data type.

Figure 51. Protein Column Selection Menu

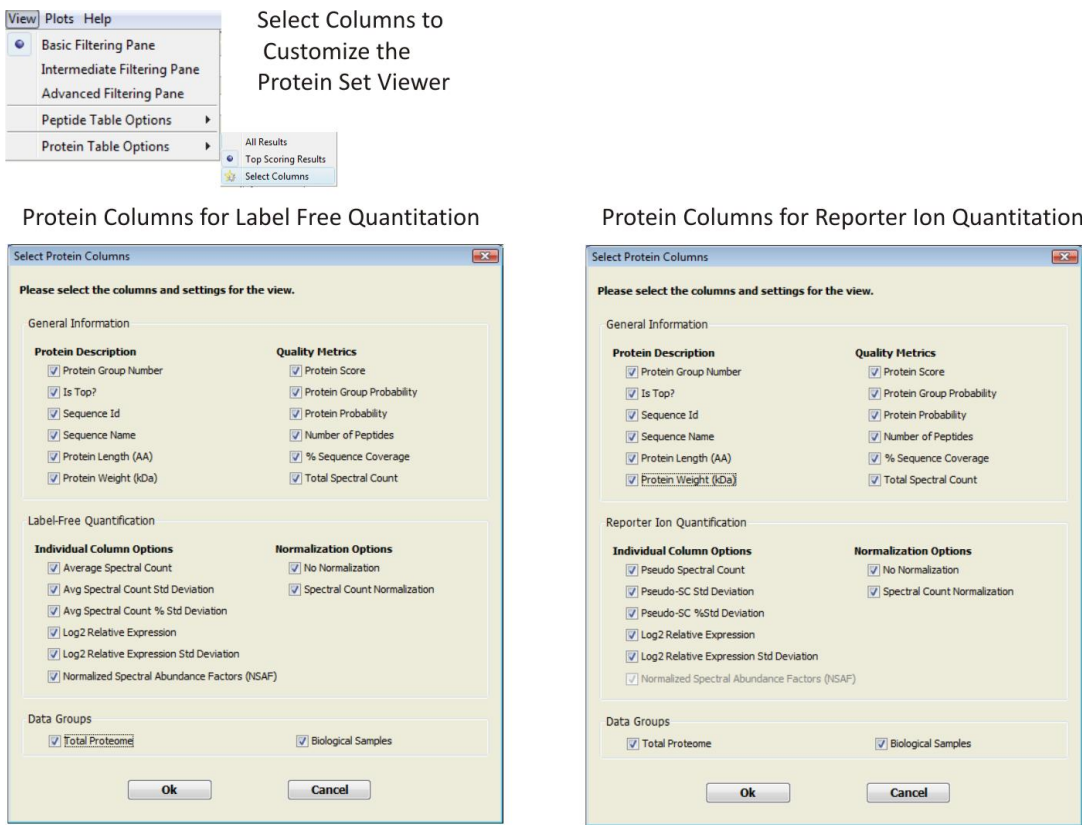


Table 23. Protein Column Selection Menu – Data Groups

Field	Description
Total Proteome	Displays the Total Values for each protein identification. Totals are calculated by summing the values across all Biological Groups and Replicates.
Biological Samples	Displays the Quality Metrics and Quantification Data for each Biological Group.

Table 24. Protein Column Selection Menu – General Information

Field	Description
Protein Group Number	Ranks the Protein Groups based on Total Score.
Is Top?	TOP indicates that the protein sequence can account for all peptides within a Protein Group. Selecting Is Top? hides the other members within the protein group. See Protein Group topic for more information.
Sequence ID	In the default database parse settings for ProteoIQ, the Sequence ID represents the string (accession number, etc.) used to access a given protein in a public repository.
Sequence Name	Description of the protein or gene in the database as defined by the Protein Name Parse Rule.
Protein Length	Number of amino acids in the protein.
Protein Weight	The sum of the molecular masses for all amino acids in the protein amino acid sequence.
Protein Group Probability	Probability that the Protein Group is identified. Protein Group Probability does not penalize proteins if they contain peptides identified in multiple proteins within a group.
Protein Probability	Probability that a protein within a Protein Group is identified. Protein Probability penalizes proteins if they contain peptides identified in multiple proteins within a group.
Number of Peptides	The number of peptides matching a protein.
% Sequence Coverage	Calculated as the number of amino acids in the identified peptides divided by the total number of amino acids in the protein sequence.
Total Spectral Count	The sum of all spectra generating peptide identifications to a given protein, i.e., the number of redundant peptide spectra.

Table 25. Protein Column Selection Menu – Label Free Quantification

Field	Description
Average Spectral Count	Calculated by dividing the Sum of the Total Spectral Counts for a protein in each Replicate by the number of Replicates.
Avg Spectral Count Std Deviation	If >2 replicates are assigned, ProteoIQ calculates the standard deviation for each protein identification using the Total Spectral Counts and Average Spectral Count in each replicate for a given Biological Group.
Avg Spectral Count %Std Deviation	The percent standard deviation (more commonly referred to as relative standard deviation, %RSD) is calculated as: (Avg Spectral Count Std Deviation / Average Spectral Count) X 100%.
Log ₂ Relative Expression	Calculated by taking the Log ₂ of the Relative Expression Value. For more detailed information see: Quantitation Overview – Label Free
Log ₂ Relative Expression Std Deviation	If >2 replicates are assigned, ProteoIQ calculates the standard deviation of Log ₂ Relative Expression Value across all replicates within a Biological Groups.
No Normalization	Calculates Relative Expression and Log ₂ Relative Expression without normalization.
Spectral Count Normalization	Total Spectral Counts are first normalized by comparing the total spectral counts for all proteins identified in each biological group. The normalized Spectral Counts are then used to calculate Relative Expression and Log ₂ Relative Expression. For more information see the Sample Normalization topic.
Protein Size Normalization	Displays the Protein Size Normalized Total Spectral Counts for each Biological Group. For more information on how the Protein Normalization is performed see the Protein Normalization topic.

Table 26. Protein Column Selection Menu – Reporter Ion Quantification

Field	Description
Pseudo Spectral Count	Calculated by normalizing the reporter ion intensities for each peptide then summing the normalized reporter ion intensities for all peptides matching a protein. For detailed information on how this calculation is performed see topic: Pseudo Spectral Count Quantitation
Pseudo Spectral Count Std Deviation	If >2 replicates are assigned, ProteoIQ calculates the standard deviation for each protein identification using the Pseudo Spectral Counts and Average Pseudo Spectral Count in each replicate for a given Biological Group.
Pseudo Spectral Count %Std Deviation	The percent standard deviation (more commonly referred to as relative standard deviation, %RSD) is calculated as: (Avg Pseudo Spectral Count Std Deviation / Average Pseudo Spectral Count) X 100%.
Log ₂ Relative Expression	Calculated by taking the Log ₂ of the Relative Expression Value. For more detailed information see: Quantitation Overview – Reporter Ion
Log ₂ Relative Expression Std Deviation	If >2 replicates are assigned, ProteoIQ calculates the standard deviation of Log ₂ Relative Expression Value across all replicates within a Biological Group.
No Normalization	Calculates Relative Expression and Log ₂ Relative Expression without normalization using the Pseudo Spectral Count.
Spectral Count Normalization	Total Pseudo Spectral Counts are first normalized by comparing the total pseudo spectral counts for all proteins identified in each biological group. The normalized Pseudo Spectral Counts are then used to calculate Relative Expression and Log ₂ Relative Expression. For more information see the Sample Normalization topic.

Customize Protein and Peptide Tables in Protein Sequence Viewer

Figure 52 shows the Protein and Peptide Column Selection Menu in the [Protein Sequence Viewer](#). The Column Selection Menu is accessible by choosing:

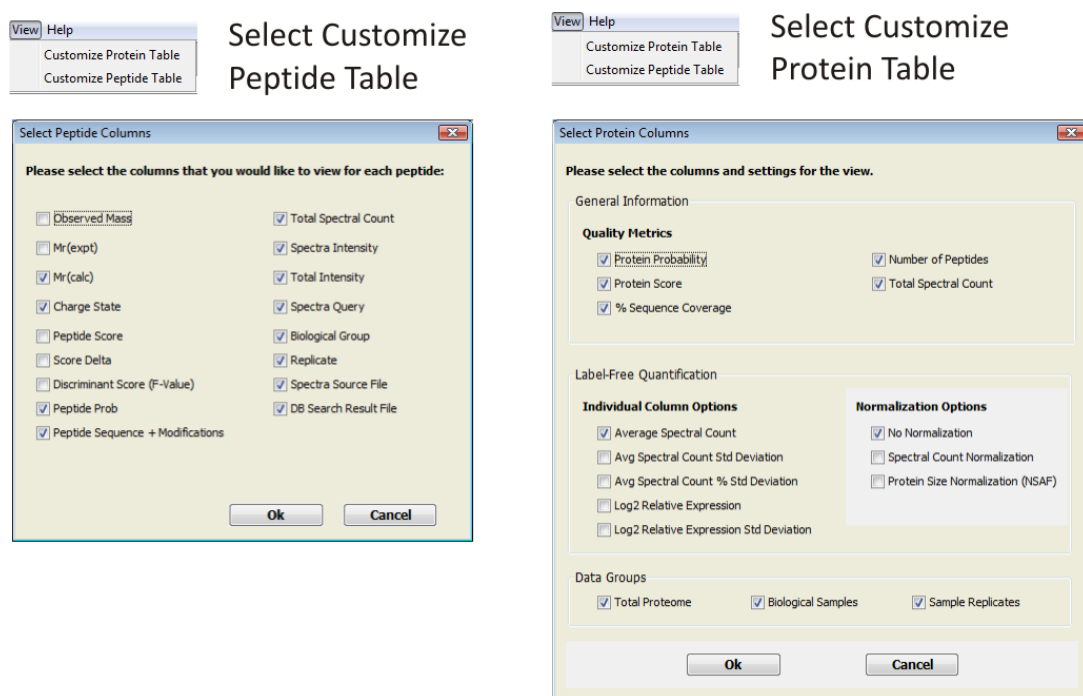
View > Customize Protein Table or View > Customize Peptide Table

Select columns to display by checking the box adjacent to each column name.

Table 22 contains descriptions of the data types in the Customize Peptide Table Menu.

Table 23, Table 24, Table 25, and Table 26 provides descriptions of each data types in the Customize Protein Table Menu.

Figure 52. Peptide and Protein Table Column Selection Menu in Protein Sequence Viewer



Note: Selecting "Sample Replicates" in the Protein Table Column Selection Menu will display the results associated with each Replicate.

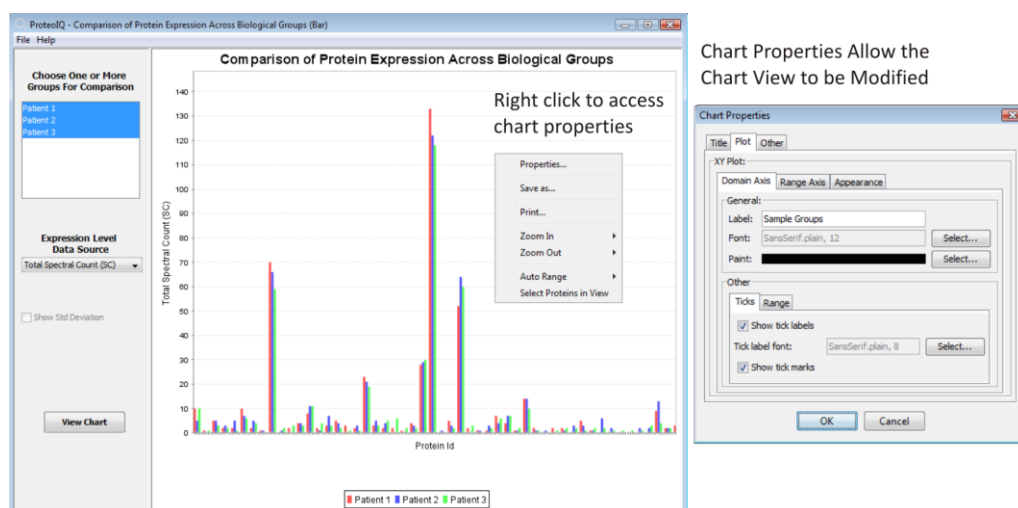
Modifying Plot Views

All of the Plot Views in ProteoIQ can be customized by right clicking in the plot area, then selecting Properties as shown in Figure 53.

Fields that can be modified include:

1. Chart Header Text
2. X-Axis Label Text
3. Y-Axis Label Text
4. Line Colors
5. Font Type and Colors
6. Axis Tick Marks
7. Plot Background Color

Figure 53. Chart Properties Menu



Select Proteins In View

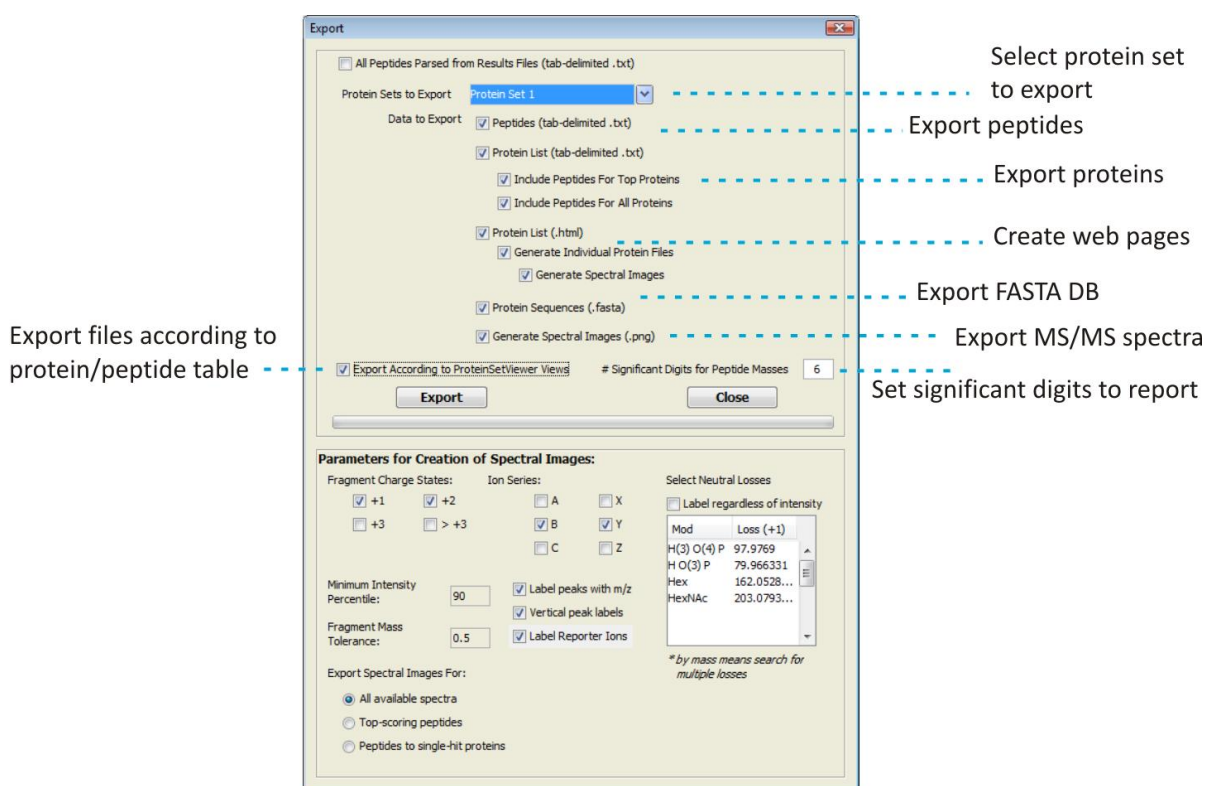
The select proteins in view function enables the selection of proteins in the protein set view that are displayed within the viewable area of each plot. This function is particularly useful for selecting proteins within the **Relative Expression Differences** and **Relative Expression Changes** expression plots.

Exporting Results

To access the Export Menu select **File > Export**. Figure 54 shows that the Export Menu is divided into three sections:

- 🔍 **Protein Sets to Export**
- 🔍 **Data to Export**
- 🔍 **Parameters for Creation of Spectral Images**

Figure 54. Export Menu



Protein Sets to Export

Select the Protein Sets to export. The Protein Sets are listed in the [Protein Set Navigation Bar](#). The check boxes below specify what **Data Formats** will be saved for each protein set selected.

Data to Export

Table 27 describes the export formats.

Table 27. Export Formats

Field	Description
All Parsed Peptides From Results Files	Saves a tab-delimited file containing all peptides extracted from the Mascot, SEQUEST, or X!Tandem results.
Peptides	Saves a tab-delimited file containing all peptides and related information from the selected protein sets.
Protein List	The Saves a tab-delimited file containing all protein identifications, scores, and related information from the selected protein sets. If Include Peptides For Top Proteins is selected, the peptide identifications associated with each protein assignment will be included in the exported file.
Protein list html	Saves an html file containing all protein and peptide identifications. If Generate Individual Protein Files is selected hyperlinked protein files are generated that contain the protein sequence coverage information.
Protein Sequences	Saves a FASTA file containing the protein sequences for all identified proteins in the selected protein sets.
Generate Spectral Images	Creates PNG image files for the selected MS/MS spectra. If selected the user must define how the spectra will be annotated using the Parameters for Creation of Spectral Images Menu.
Export According to ProteinSetViewer Views	Selecting Export According to ProteinSetViewer Views will create all tab-delimited and web page exports using the column and row settings of the selected protein or peptide set.
# Significant Digits for Peptide Masses	Specifies the number of decimal places to report for the Peptide Mr(expt), Mr(calc), and Mr (Observed) in the Peptides (tab-delimited.txt) export.

Parameters for Creation of Spectral Images

When Generate Spectral Images is selected the user must define how the spectra will be annotated using the Parameters for Creation of Spectral Images Menu. Table 28 describes the parameters in the Spectral Images Menu.

Table 28. Spectral Images Menu

Field	Description
Fragment Charge States	Four selections are available, +1, +2, +3, and > +3. Selection of a charge state will label ions in the MS/MS spectrum that arise from the selected ion series (y, b, etc.) at the assigned charge state. For example if a charge state of +2 and an ion series of y are selected, ProteoIQ will label the peaks in the MS/MS spectrum that correlate to the m/z of predicted, doubly charged y ions from the identified peptide. By default, charge is considered $M+H^+$, $M+H_2^{++}$, and $M+H_3^{+++}$, for +1, +2, and +3 etc.
Ion Series	From the peptide sequence, ProteoIQ predicts the m/z of the common fragment ions, y, b, a, c, x, and z. Selecting an ion series labels the MS/MS spectrum if there are ions present (of the selected ion series) that meet the criteria set in the left panel.
Minimum Intensity Percentile	ProteoIQ generates a distribution of intensities for the fragment ions present in each MS/MS spectrum. The intensities are binned to create a histogram based on the frequency of intensities observed in the MS/MS spectrum. The minimum intensity percentile requires that the peaks in the MS/MS spectrum must be above the setting in order to be labeled with an ion series annotation. For example, a minimum intensity percentile of 90% requires that a peak in the MS/MS spectrum has intensity greater than 90% of all ions in the spectrum to be considered for annotation with an ion series label. Lowering the minimum intensity percentile allows ions of lesser intensity to be considered for ion series annotation.
Fragment Mass Tolerance	Specifies the error window for labeling of fragment ions with ion series annotations. Units are Da.

Field	Description
Label Reporter Ions	Adds the reporter ion label to the reporter ions in the MS/MS spectrum.
Label Peaks with Mass to Charge	Adds the m/z of the ion along with the ion series label.
Selected Neutral Losses	In the Selected Neutral Losses table, masses of expected neutral losses can be entered, and then selected, for annotation in MS/MS spectrum. The mass of the neutral loss will be adjusted based on the charge of the observed precursor ion and subtracted from the precursor m/z. For example, if the neutral loss mass is entered as -98 Da and the observed precursor ion has a charge state of $M+H_2^{++}$ then ProteoIQ will search for an ion in the MS/MS spectrum, over the minimum intensity percentile that is 49 m/z units below the m/z of the observed precursor ion.
Label Regardless of Intensity	Selecting Label Regardless of Intensity removes the minimum intensity percentile filter when labeling ions resulting from the selected neutral losses.
All Available Spectra	Exports spectral images for ALL spectra in the selected protein set.
Top-Scoring Peptides	Exports spectral images for ONLY the top scoring peptides in the selected protein set. Only one spectral image is created per peptide.
Peptides to Single-Hit Proteins	Exports spectral images for peptides present on proteins with ONLY a single peptides assignment.

Evaluating Results

The following sections provide an overview on how to interpret your results in ProteoIQ. The following contents are covered in this section.

- 🔍 [Peptide Identification](#)
- 🔍 [Protein Identification](#)
- 🔍 [Whole Proteome Visualization](#)
- 🔍 [Quantitation](#)

Peptide Identification

In ProteoIQ there are three methods to assist in establishing a threshold of confidence for peptide identifications generated from database search results. These include traditional methods of [Filtering by Scores](#) or related metrics (Score Delta, Charge, etc.), [Peptide Probabilities](#), and [Peptide False Discovery Rates](#). Manual validation of individual MS/MS spectra is also supported in ProteoIQ.

- 🔍 [Score Filters](#)
- 🔍 [Peptide Probability](#)
- 🔍 [Peptide False Discovery Rate](#)
- 🔍 [Validating MS/MS Spectra](#)

Note: Peptide probability and False Discovery Rate estimations are HIGHLY dependent on the size of the data set. It is always recommended that you evaluate the probability and false discovery rate plots to determine if these calculations are valid.

Score Filters

For data sets with fewer than 500 peptides, native database search scores should be used to filter peptide identifications. Setting thresholds for confident peptide identification can be performed using native **Mascot Ion Scores**, **SEQUEST Xcorr's**, and **!Tandem Hyper Scores**. Two methods can be used to remove peptides by score as shown in Figure 55.

Figure 55. Score Filters

Combining Peptide Scores with Charge State Filters

Sort by Score

Copy Selected to New Protein Set

Validating peptide identifications by Score in the Filter Pane

1. Enter **Min Peptide Score** in **Filter Pane**
2. Select **Create New Protein Set**

Validating peptide identifications by Score in the Peptide Set Viewer

1. Sort peptides by **Score** by selecting the Score column header
2. Highlight to select peptides of interest
3. Right click and select **Copy Selected to New Protein Set**

Combining Score Filters with Charge State

A common approach for validating peptide identifications is to combine Score Filters with Charge State Filters. The assumption is that as charge increases the size of the peptide will also increase. Since scores are directly related to the number of matched fragment ions then different score thresholds should be used for peptides of different size.

To perform this analysis in ProteoIQ, simply enter the Score in Min Peptide Score and then select the Charge State in the Filter Pane.

Peptide Probabilities

ProteoIQ calculates probabilities that estimate the likelihood that the peptide assignment to a given MS/MS spectra is correct. Probability thresholds can be applied using the Filter Pane or by Sorting peptides by probability in the Peptide Set Viewer then copying the selecting peptides to a new protein set as shown in Figure 55.

Validating peptide identifications by probability in Filter Pane

1. Enter **Min Peptide Probability** in **Filter Pane**
2. Select **Create New Protein Set**

Validating peptide identifications by Peptide Probability in Peptide Set Viewer

1. Sort peptides by **Probability** by selecting the **Prob** column header
2. Highlight to select peptides of interest
3. Right click and select **Copy Selected to New Protein Set**

If probability thresholds are utilized it is critically important to evaluate the distributions used to calculate the probability values.

Assessing the Quality of Probability Calculations

Discriminant Score (F-value) Plots (Figure 56) should be used to evaluate the legitimacy of the probability calculations. To access the Discriminant Score Plots select: **Plots > Proteome Validation > Peptide Probability**. To view distributions for peptides of a given charge state Select the Charge State of interest in the Select The Plot to View: drop down menu.

Figure 56. Peptide Probability Plot

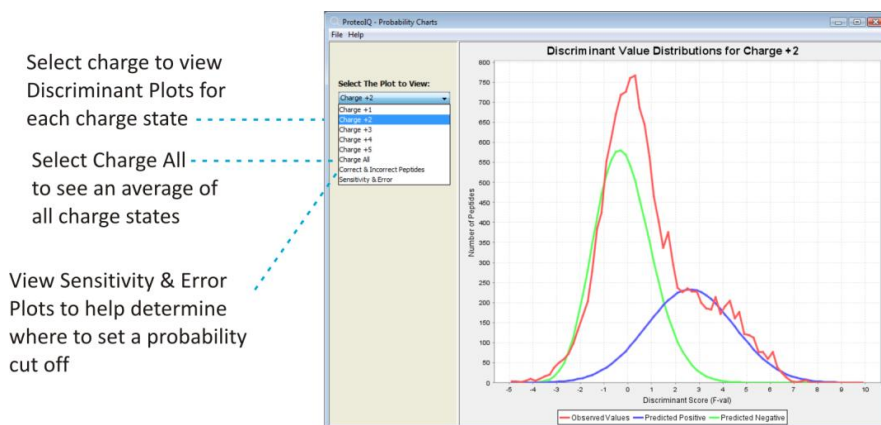
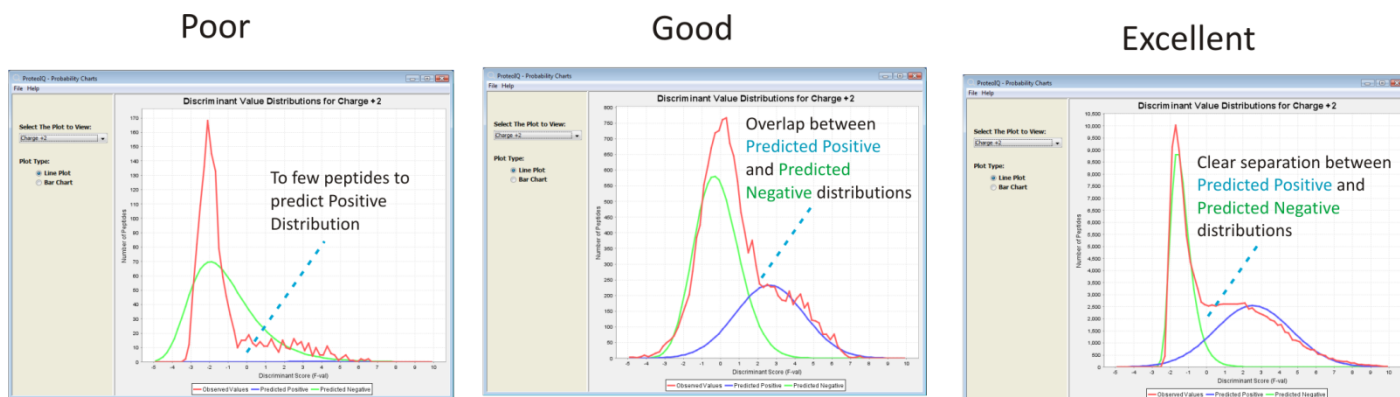


Figure 57 shows an example of how the Discriminant Score Distributions should be used to evaluate if probability based peptide identification is reliable for your data set. The Discriminant Score Distribution Plot contains three data series:

1. Distribution of experimental discriminant score – **RED**
2. Predicted distribution of negative peptide assignments - **GREEN**
3. Predicted distribution of positive peptide assignments – **BLUE**

For the peptide probabilities to be accurate the plot should show a good separation between the predicted positive and predicted negative distributions and both distributions should fit well with the distribution of experimental peptide discriminant scores.

Figure 57. Example Discriminant Score Distributions



Peptide False Discovery Rates

Peptide False Discovery Rates (pep-FDR) can be calculated and utilized to filter large peptide sets in ProteoIQ. To perform pep-FDR you will need to have searched your MS/MS spectra against a target and decoy database prior to running ProteoIQ.

The following steps should be followed to perform and pep-FDR analysis.

1. Load target and decoy database searches in **Select Files for Parsing** window of the **New Project Wizard**.
2. In the Peptide Parsing Options enter the following parameters:
 - a. **Min Score** = 0
 - b. **Min Peptide Length** = 5
 - c. **Min Peptide Prob** = 0
 - d. **Min Protein Prob** = 0
 - e. **Max Protein FDR** = 10% (Note: This value will be adjusted once the protein set is created.)
3. Once the protein set is created, select **View > Intermediate Filtering Pane**
4. Set **Min Peptide FDR %** = 1.0 and leave all other parameters blank (figure 56).
5. Select Score Metric from **Metric to Use for FDR** drop down menu.
6. Select **No Filter** in the **Filter Previous Set** drop down menu.
7. Click Create New Protein Set

Figure 58. Setting Peptide FDR in Filtering Pane

False-Discovery Rate Filters	
Metric to Use for FDR	Mascot ions score
Max Peptide FDR (%)	1.0
Max Protein FDR (%)	
Starting Pep Coverage for ProFDR	
Probability Filters	
Min Peptide Probability	
Min Protein Probability	
Min Protein Group Probability	

Assessing the Quality of Peptide False Discovery Rate Calculations

Peptide False Discovery Rate Plots (Figure 59) should be used to evaluate the legitimacy of the peptide false discovery rate calculations. To access the **Peptide False Discovery Rate Plot** select: **Plots > Proteome Validation > Peptide FDR**. To view distributions of peptides from target and decoy database search results for different Score Metrics, select the Score Metric from the Select the Plot to View: drop down menu.

Figure 59. Peptide False Discovery Rate Plot

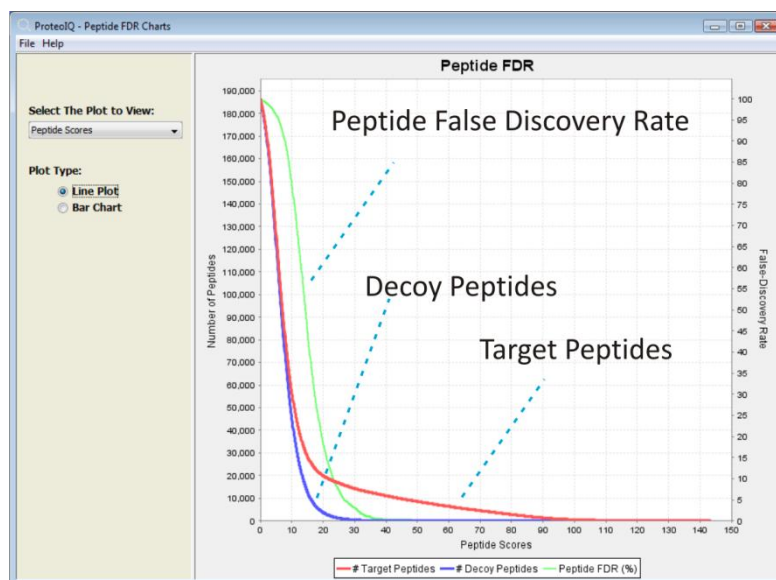


Figure 59 shows an example of how the Peptide FDR Plot should be used to evaluate if Peptide False Discovery Rate based peptide identification is reliable for your data set. The Peptide FDR Plot contains three data series:

1. Distribution of target peptides at each score – **RED**
2. Calculated Peptide False Discovery Rate at each score threshold – **GREEN**
3. Distribution of decoy peptides at each score threshold – **BLUE**

For the Peptide FDR's to be accurate the distributions should approximate a Gaussian distribution. There should also be good separation between the target and decoy peptide distributions (Figure 59).

Validating MS/MS Spectra

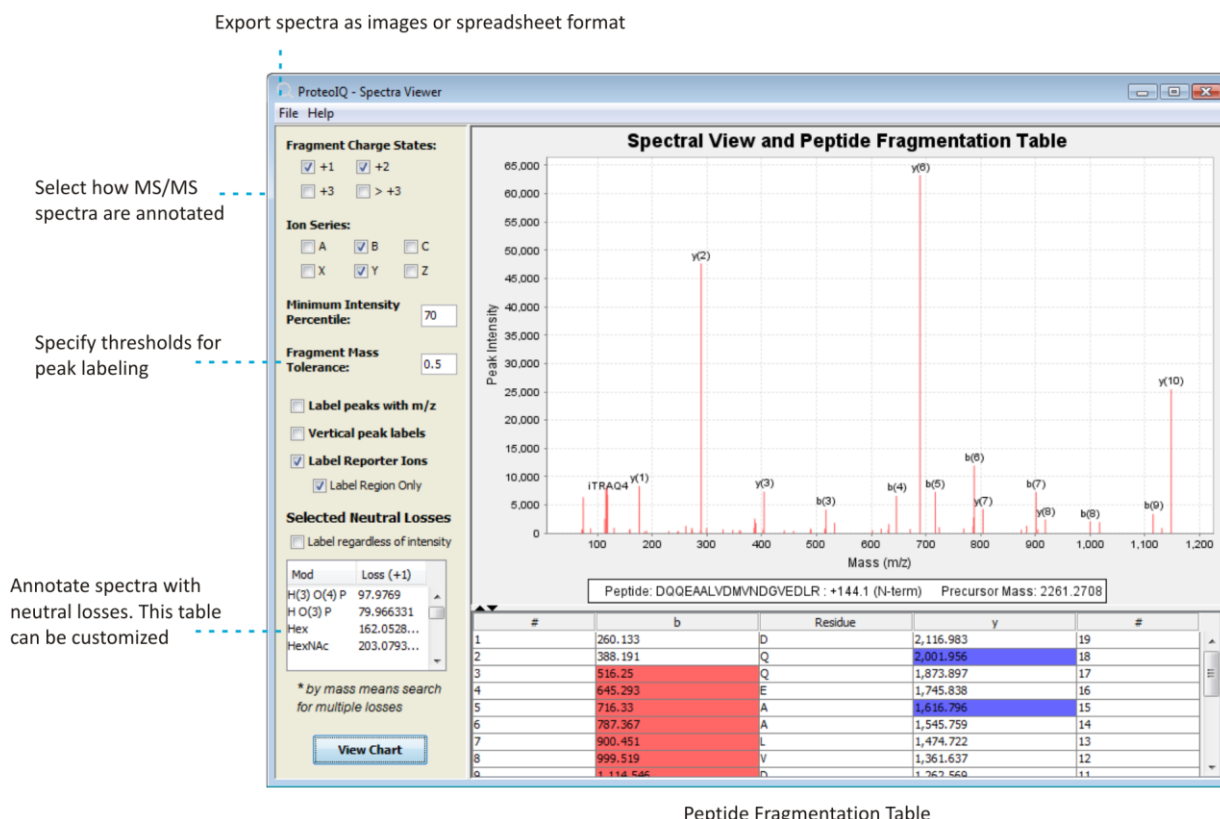
To view the MS/MS spectral assignment for any peptide identification, simply double click on the peptide sequence in the [Peptide Set Viewer](#) or [Protein Sequence Viewer](#). The following sections describe how to use the Spectral Viewer.

- 🔍 [Spectral Viewer Overview](#)
- 🔍 [Adjusting peak labeling](#)
- 🔍 [Mapping Neutral Losses](#)
- 🔍 [Validating Reporter Ion Quantitation](#)

Spectral Viewer Overview

The spectral viewer contains three panels as shown in Figure 60. The left most panel is for selection of spectral labeling parameters that are rendered in the annotated MS/MS spectrum. The bottom panel contains a table with the predicted m/z of the selected ion series. The color codes indicate that the specified ion series was detected in the selected MS/MS spectrum.

Figure 60. MS/MS Spectral Viewer



Adjusting Peak Labeling

To adjust labeling of the MS/MS spectrum, select parameters from the left panel then click **View Chart**. The spectrum and ion series table will automatically be updated if there are ions in the MS/MS spectrum that meet the criteria set in the left panel. Table 29 describes the parameters.

Table 29. Peak labeling parameters in MS/MS spectral viewer

Field	Description
Fragment Charge States	Four selections are available, +1, +2, +3, and > +3. Selection of a charge state will label ions in the MS/MS spectrum that arise from the selected ion series (y, b, etc.) at the assigned charge state. For example if a charge state of +2 and an ion series of y are selected, ProteoIQ will label the peaks in the MS/MS spectrum that correlate to the m/z of predicted, doubly charged y ions from the identified peptide. By default, charge is considered $M+H^+$, $M+H_2^{++}$, and $M+H_3^{+++}$, for +1, +2, and +3 etc.
Ion Series	From the peptide sequence, ProteoIQ predicts the m/z of the common fragment ions, y, b, a, c, x, and z. Selecting an ion series labels the MS/MS spectrum if there are ions present (of the selected ion series) that meet the criteria set in the left panel.
Minimum Intensity Percentile	ProteoIQ generates a distribution of intensities for the fragment ions present in each MS/MS spectrum. The intensities are binned to create a histogram based on the frequency of intensities observed in the MS/MS spectrum. The minimum intensity percentile requires that the peaks in the MS/MS spectrum must be above the setting in order to be labeled with an ion series annotation. For example, a minimum intensity percentile of 90% requires that a peak in the MS/MS spectrum has intensity greater than 90% of all ions in the spectrum to be considered for annotation with an ion series label. Lowering the minimum intensity percentile allows ions of lesser intensity to be considered for ion series annotation.
Fragment Mass Tolerance	Specifies the error window for labeling of fragment ions with ion series annotations. Units are Da.

Field	Description
Label Reporter Ions	Adds the reporter ion label to the reporter ions in the MS/MS spectrum. Selecting "Label Region Only" adds only a single label to the entire reporter region.
Label Peaks with Mass to Charge	Adds the m/z of the ion along with the ion series label.
Selected Neutral Losses	In the Selected Neutral Losses table, masses of expected neutral losses can be entered, and then selected, for annotation in MS/MS spectrum. The mass of the neutral loss will be adjusted based on the charge of the observed precursor ion and subtracted from the precursor m/z. For example, if the neutral loss mass is entered as -98 Da and the observed precursor ion has a charge state of $M+H_2^{++}$ then ProteoIQ will search for an ion in the MS/MS spectrum, over the minimum intensity percentile that is 49 m/z units below the m/z of the observed precursor ion.
Label Regardless of Intensity	Selecting Label Regardless of Intensity removes the minimum intensity percentile filter when labeling ions resulting from the selected neutral losses.

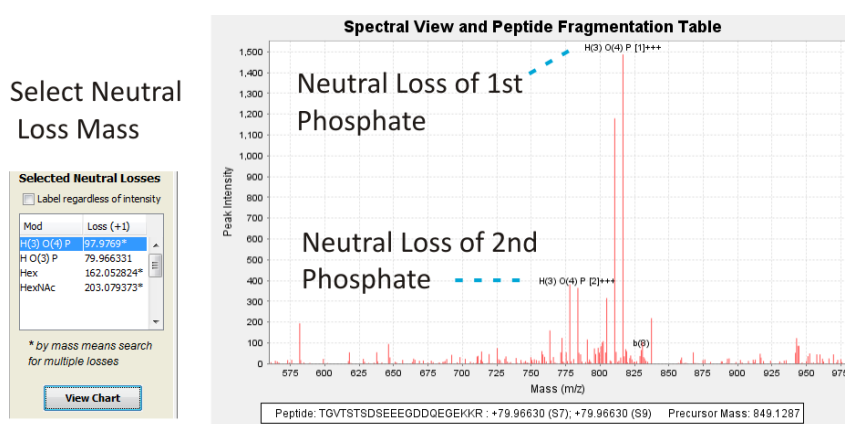
Mapping Neutral Losses

Several types of post-translational or chemical modification produce intense ions in the MS/MS spectrum because of fragmentation mechanism occurring from the loss of a stable neutral molecule. Searching for ions resulting from a neutral loss in a peptide MS/MS spectrum is supported via the Selected Neutral Loss table as shown in Figure 61. In the Selected Neutral Losses table, masses of expected neutral losses can be entered, and then selected, for annotation in MS/MS spectrum. The mass of the neutral loss will be adjusted based on the charge of the observed precursor ion and subtracted from the precursor m/z .

To search for neutral losses in the displayed spectrum:

1. If using a default neutral loss, highlight the neutral loss name and mass in the **Selected Neutral Losses** table
2. If the desired Neutral Loss mass is not listed, enter the name and the neutral mass (Mr) at the bottom of the table.
3. Select **View Chart**

Figure 61. Neutral Loss Mapping in the MS/MS Spectra View



Searching for Multiple Neutral Losses

If more than one unique neutral loss is highlighted in the Selected Neutral Losses table, ProteoIQ will search for each neutral loss based on the mass value entered in the table. In cases where multiple neutral losses of the same type occur from a single peptide these can be identified by adding a * after the neutral loss mass. For example, in the figure above, the peptide has two sites of phosphorylation. In the Selected Neutral Losses table **H(3)O(4)P, 97.9*** is entered and the MS/MS spectrum is labeled with two sequential phosphate neutral losses, labeled **H(3)O(4)P [1]** and **H(3)O(4)P [2]**.

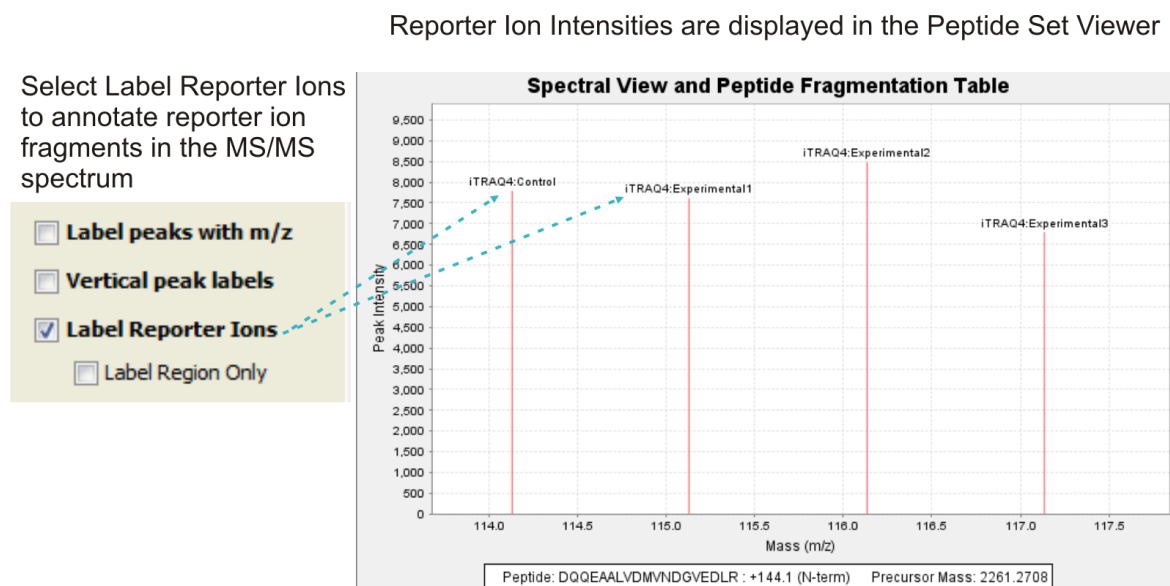
Validating Reporter Ion Quantitation

Searching for fragment ions resulting from the cleavage of an isobaric tag in a peptide MS/MS spectrum is supported via the Label Reporter Ions check box, as shown in Figure 62. Two options are available for labeling reporter ions. The Label Reporter Ions option annotates each reporter ion with the name for the biological group. When Label Region Only is selected the entire reporter region is labeled with the name of the isobaric tag used for the experiment.

To validate the quality of the reporter fragments:

1. Open the spectral viewer by double clicking on a peptide of interest
2. Click the checkbox adjacent Label Reporter Ions in the spectral viewer
3. Zoom into the reporter region (typically from 110 – 140 m/z)
4. The intensities for each reporter ion will be displayed in the Peptide Set Viewer

Figure 62. Reporter Ion validation in the MS/MS Spectral View



Protein Identification

In ProteoIQ there are three methods to assist in establishing a threshold of confidence for protein identifications generated from database search results. These include traditional methods of **Filtering by Scores** or related metrics (Score Delta, Charge, etc.), **Protein Probabilities**, and **Protein False Discovery Rates**.

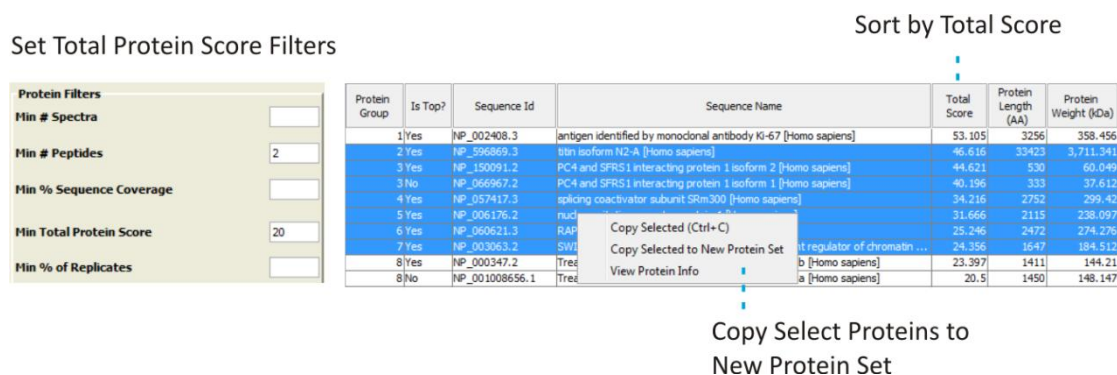
- 🔍 **Protein Score Filters**
- 🔍 **Protein Probability**
- 🔍 **Protein False Discovery Rate**

Note: Protein probability and False Discovery Rate estimations are HIGHLY dependent on the size of the data set. It is always recommended that you evaluate the probability and false discovery rate plots to determine if these calculations are valid.

Protein Score Filters

ProteoIQ sums the Peptide Scores for all peptides matching to a protein to create a Total Score. Two methods can be used to filter proteins by Total Score as shown in Figure 63.

Figure 63. Protein Score Filters



Validating protein identifications by Total Score in the Filter Pane

1. Enter **Min Total Protein Score** in **Filter Pane**
2. Select **Create New Protein Set**

Validating protein identifications by **Total Score** in the Protein Set Viewer

1. Sort peptides by **Score** by selecting the **Total Score** column header
2. Highlight to select proteins of interest
3. Right click and select **Copy Selected to New Protein Set**

Protein Probability

ProteoIQ calculates probabilities that estimate the likelihood that the protein assignment to be correct based on the peptide probabilities for the identified peptides mapping to the protein. Probability thresholds can be applied using the [Filter Pane](#) or by sorting proteins by probability in the [Protein Set Viewer](#) then copying the selecting proteins to a new protein set as shown in Figure 8.

Validating protein identifications by probability in the Filter Pane

1. Enter [Min Protein Probability](#) in [Filter Pane](#)
2. Select [Create New Protein Set](#)

Validating protein identifications by Protein Probability in Protein Set Viewer

1. Sort proteins by [Probability](#) by selecting the [Prob](#) column header
2. Highlight to select proteins of interest
3. Right click and select [Copy Selected to New Protein Set](#)

If probability thresholds are utilized it is critically important to evaluate the distributions used to calculate the probability values. For more details, see the [Assessing the Quality of Probability Calculations](#) section.

Protein False Discovery Rate

Protein False Discovery Rates (pro-FDR) can be calculated and utilized to filter large protein sets in ProteoIQ. To perform pro-FDR you will need to have searched your MS/MS spectra against a target and decoy database prior to running ProteoIQ.

The following steps should be followed to perform a pro-FDR analysis.

1. Load target and decoy database searches in **Select Files for Parsing** window of the New Project Wizard.
2. In the Peptide Parsing Options enter the following parameters:
 - a. **Min Score** = 0
 - b. **Min Peptide Length** = 5
 - c. **Min Peptide Prob** = 0
 - d. **Min Protein Prob** = 0
 - e. **Max Protein FDR** = 1% (Note: This value is dependent upon your desired level of stringency and can be adjusted once the protein set is created.)

The following steps should be followed to perform a pro-FDR analysis using the Filter Pane (Figure 58).

1. Once the protein set is created, select **View > Intermediate Filtering Pane**
2. Set **Min Protein FDR %** = 1.0 and leave all other parameters blank (Figure 58).
3. Select Score Metric from **Metric to Use for FDR** drop down menu.
4. Set the **Starting Pep Coverage** to 3.
5. Select **No Filter** in the **Filter Previous Set** drop down menu.
6. Click Create New Protein Set

Assessing the Quality of Protein False Discovery Rate Calculations

Protein False Discovery Rate Plots and Output Notes Pane (Figure 64) should be used to evaluate the legitimacy of the protein false discovery rate calculations. To access the Protein False Discovery Rate Plot select: **Plots > Protein FDR**. To view distributions of proteins from target and decoy database search results for different Peptide Coverage Levels, select the Peptide Coverage from the Select the Plot to View: drop down menu.

Note: When evaluating a Pro-FDR in the Protein FDR Plot it is important to use the Maximum Starting Peptide Coverage Level to assess the distributions. Since ProteoIQ does not calculate Pro-FDRs at EVERY score threshold the distributions will not appear as smooth as in the Pep-FDR and Probability Plots. Cross referencing with the Output Notes Pane will aid in evaluating the validity of the Pro-FDR calculations.

Figure 64. Protein False Discovery Rate Plot

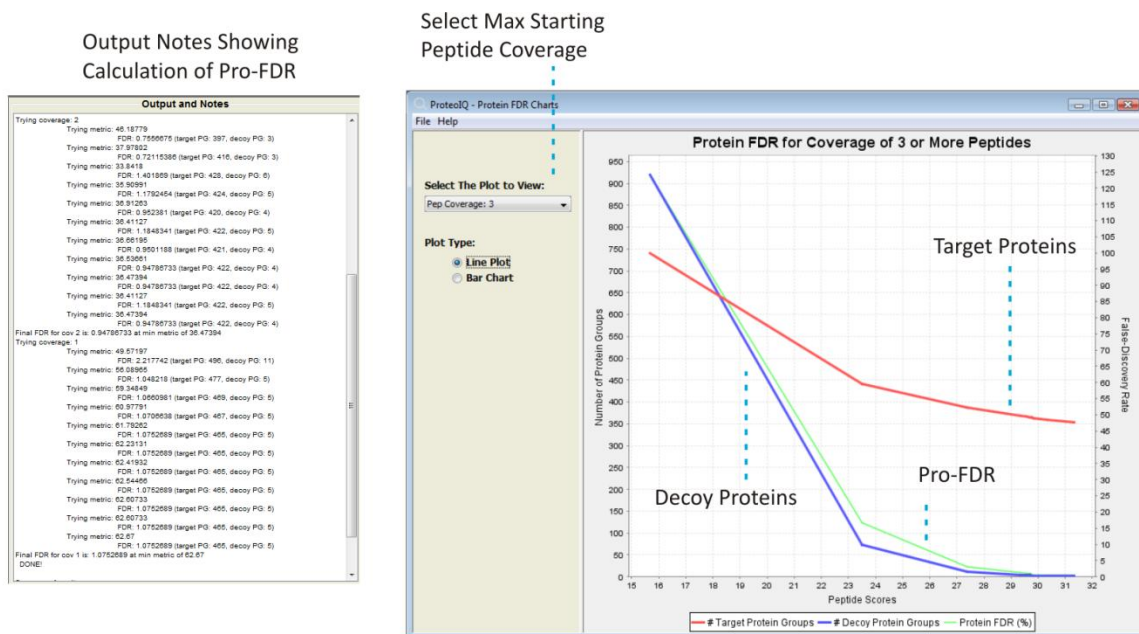


Figure 64 shows an example of how the Protein-FDR Plot should be used to evaluate if Protein False Discovery Rate based protein identification is reliable for your data set. The Protein-FDR Plot contains three data series:

1. Distribution of target proteins at each score – **RED**
2. Calculated Protein False Discovery Rate at each score threshold – **GREEN**
3. Distribution of decoy proteins at each score threshold – **BLUE**

For the Protein FDR's to be accurate the distributions should approximate a Gaussian distribution. There should also be good separation between the target and decoy peptide distributions (Figure 64). It is important to view distributions based on the Maximum **Starting Peptide Coverage** since this plot will contain the largest number of data points for viewing a complete distribution. It is also helpful to view the Output Notes section to view the number of proteins at each score threshold for calculation of the **Protein False Discovery Rates**.

Whole Proteome Visualization

To view a snapshot of the entire protein set, select the following:

- 🔍 [Pie Chart](#)
- 🔍 [Venn Diagram – Group Summary](#)
- 🔍 [2D Gel View](#)
- 🔍 [1D Gel View](#)
- 🔍 [1D Gel View by Protein](#)

Access [Pie Chart](#): **Plots > Experimental Reproducibility > Pie Chart**

Access [Venn Diagram](#): **Plots>Comparative Proteomics > Venn Group Summary (Venn)**

Access [2D Gel View](#): **Plots > Gel Based Proteomics > 2D Gel View**

Access [1D Gel View](#): **Plots > Gel Based Proteomics > 1D Gel View**

Access [1D Gel View by Protein](#): **Plots > Gel Based Proteomics > Mol Wt Cluster Plot**

Pie Chart

The Pie Chart (Figure 65) presents an overview of the entire proteome compared across Biological Groups. To view the Pie Chart select: **Plots > Experimental Reproducibility > Pie Chart**. Table 30 describes the metrics that can be displayed.

Figure 65. Pie Chart

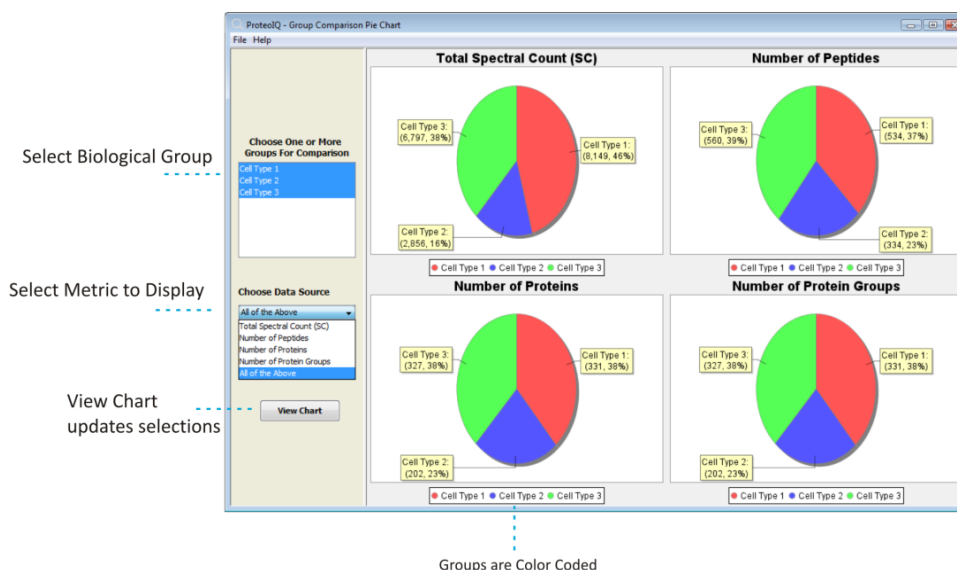


Table 30. Pie Chart Parameters

Field	Description
Total Spectral Count	Compares the Total Spectral Counts for all proteins between Biological Groups. The Total Spectral Count for a group is used to perform Sample Normalization for spectral counting. Viewing the Total Spectral Count Pie Chart can be used to determine if a certain group was over or under sampled.
Number of Peptides	Compares the number of peptides for all proteins between Biological Groups.
Number of Proteins	Compares the total number of proteins between Biological Groups.
Number of Protein Groups	Compares the total number of protein groups between Biological Groups.
All of the Above	Generates four pie charts showing all of the above metrics.

Venn Diagram - Group Summary (Venn)

The Venn Diagram (Figure 66) displays the relationship between selected biological groups in terms of protein groups, proteins or peptides. To view the Venn Diagram select: **Plots > Comparative Proteomics > Venn Group Summary (Venn)**. Table 31 describes the metrics that can be displayed and differences between the **Traditional** and **Strict** Venn Diagrams are discussed in the following sections.

- 🔍 [Traditional Venn Diagram Labeling](#)
- 🔍 [Strict Venn Diagram Labeling](#)

Figure 66. Group Summary Venn Diagram

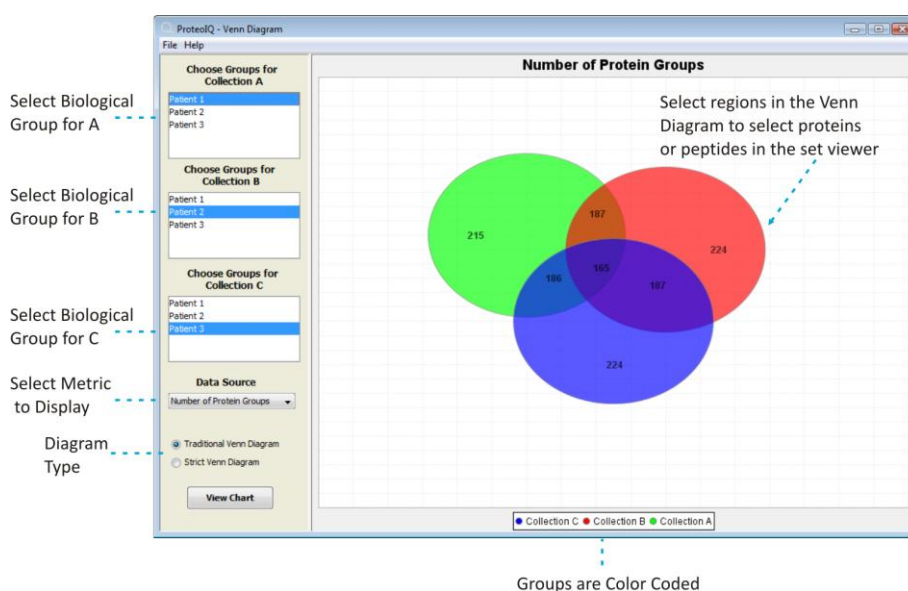


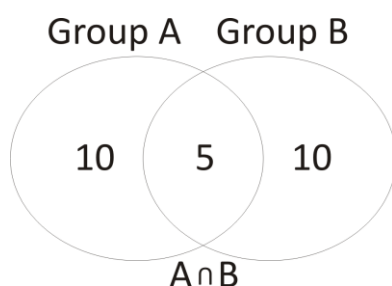
Table 31. Group Summary Venn Diagram Data Sources

Field	Description
Number of Peptides	Compares the number of peptides for all proteins between Biological Groups.
Number of Proteins	Compares the total number of proteins between Biological Groups.
Number of Protein Groups	Compares the total number of protein groups between Biological Groups.

Traditional Venn Diagram Labeling

Under traditional labeling the total number of protein groups, proteins or peptides are reported for each biological group along with the intersection data across biological groups. Under Traditional Labeling there is no concept of uniqueness as indicated in the example below:

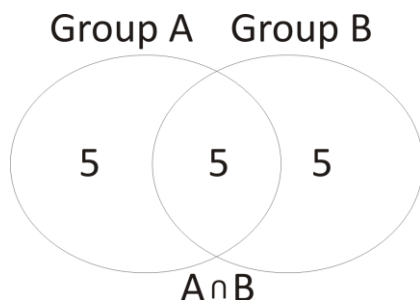
	Biological Group A	Biological Group B	Intersection A and B
Number of proteins	10	10	5



Strict Venn Diagram Labeling

Under strict labeling the non-intersection areas display the protein groups, proteins, or peptides that are unique to each biological group as displayed in the example below:

	Biological Group A	Biological Group B	Intersection A and B
Number of proteins	10	10	5



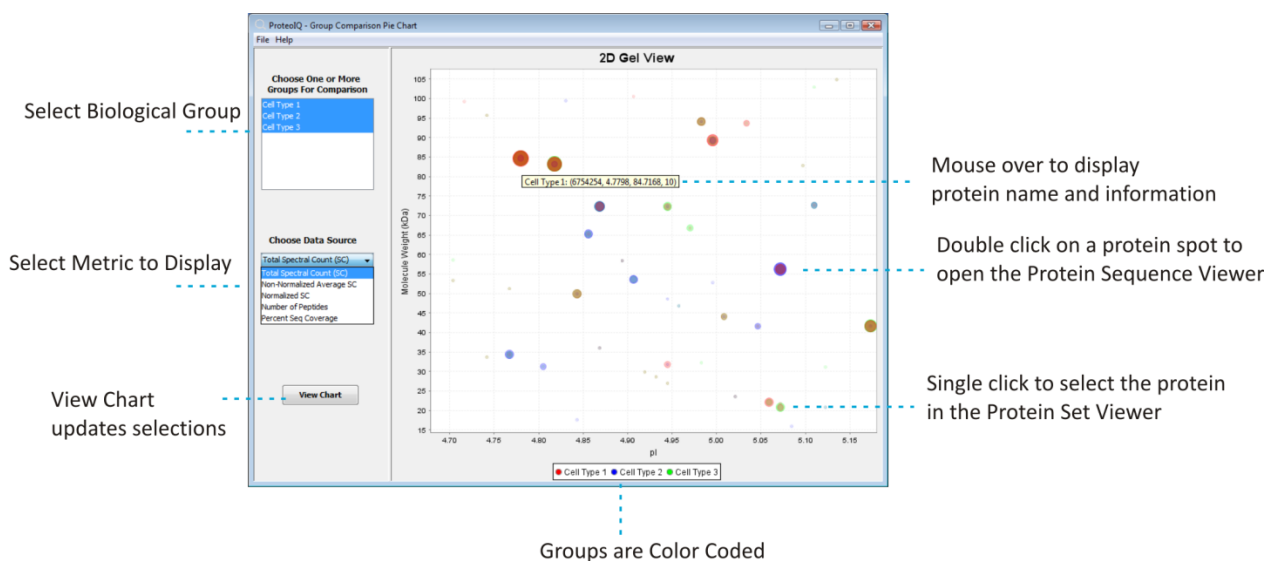
2D Gel View

The 2D Gel View (Figure 67) presents an overview of protein expression across the selected biological groups. To view the 2D Gel View select: **Plots > Gel Based Proteomics > 2D Gel View**.

To display a 2D Gel View:

1. Select the biological groups to display from the Choose One or More Groups for Comparison menu.
2. Choose the metric to display (Note: the size of darkness of the spot is dependent upon the selected value).
3. Click View Chart

Figure 67. 2D Gel View



Interacting With The 2D Gel View

In the 2D Gel View the Y-axis is theoretical molecular weight and the X-axis is the theoretical isoelectric point. Both values are calculated from the properties of the amino acids in the protein and do not include contributions from detected modifications.

The size of the protein spot is directly proportional to the data source value. For example, if Total Spectral Counts is selected then a protein with 200 spectral counts will have a spot that is twice the size and darkness as a protein with 100 spectral counts.

Note: Single clicking on a protein spot will select the protein in the protein set viewer. Double clicking will open the protein sequence viewer.

1D Gel View

The 1D Gel View (Figure 68) is used to display the results of a 1D GeLC analysis. The following sections discuss how to setup up a ProteoIQ project for use with the 1D Gel View.

- 🔍 [Experimental Approach for GeLC Analysis](#)
- 🔍 [Preparing Results for GeLC Analysis in ProteoIQ](#)
- 🔍 [How to Display a 1D Gel View](#)
- 🔍 [Interpreting the 1D Gel View](#)

Experimental Approach for GeLC Analysis

The underlying assumption when using this plot is that the following experimental approach has been used:

1. Proteins were separated on a 1D Gel.
2. The gel lane was sliced into sections.
3. Proteins in each gel slice were in-gel digested to produce peptides.
4. The peptides from each gel slice were independently analyzed by LC-MS/MS.
5. Each MS/MS file generated one or more database search results.

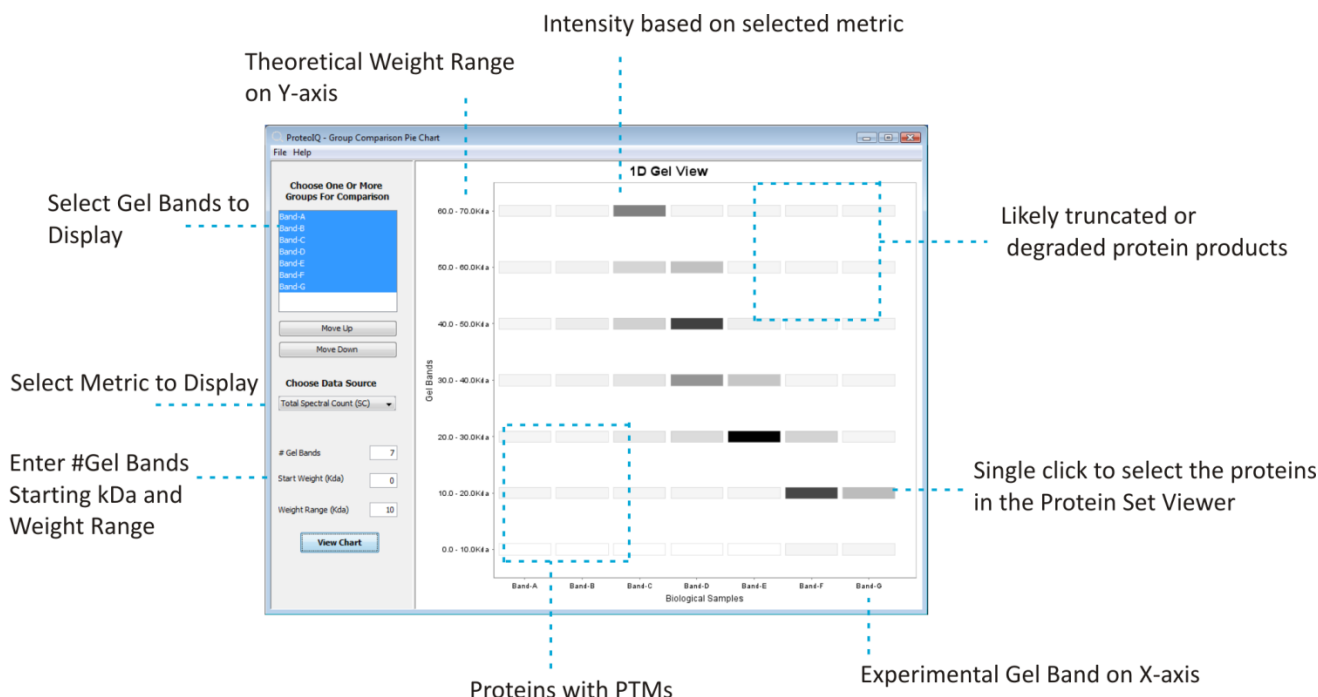
Preparing results for GeLC Analysis in ProteoIQ

1. Set the number of Biological Samples to equal the number of gel slices in the [Describe Proteome](#) window of the [New Project Wizard](#).
2. Name to Biological Samples according to gel slice. It is recommended that you number the groups using strict alpha-numeric ordering (01, 02.....10).
3. Group the database search results by gel slice in the [Select Files for Parsing](#) window.
4. Once the protein set has been created, open the 1D Gel View by selecting: **Plots > Gel Based Proteomics > 1D Gel View**.

How to display a 1D Gel View

1. Select the biological samples (Gel Slices) to display from the Choose One or More Groups for Comparison menu.
2. Choose the metric to display (Note: the darkness of the cell is dependent upon the selected value).
3. Click View Chart

Figure 68. 1D Gel View



Interpreting the 1D Gel View

When gel slices (Biological Samples) are selected they appear on the X-axis. The theoretical gel bands (created by selecting the #Gel Bands and Weight Range) are displayed on the Y-axis. Each cell is then color coded by the amount of the metric chosen in the Choose Data Source drop down menu.

In the example shown in Figure 68, Total Spectral Counts is selected. For each cell, the spectral counts for all proteins appearing in the theoretical molecular weight range and experimental gel slice are summed. The darker the cell, the more the spectral counts.

By selecting a cell, ALL of the proteins appearing in that cell are selected in the Protein Set Viewer.

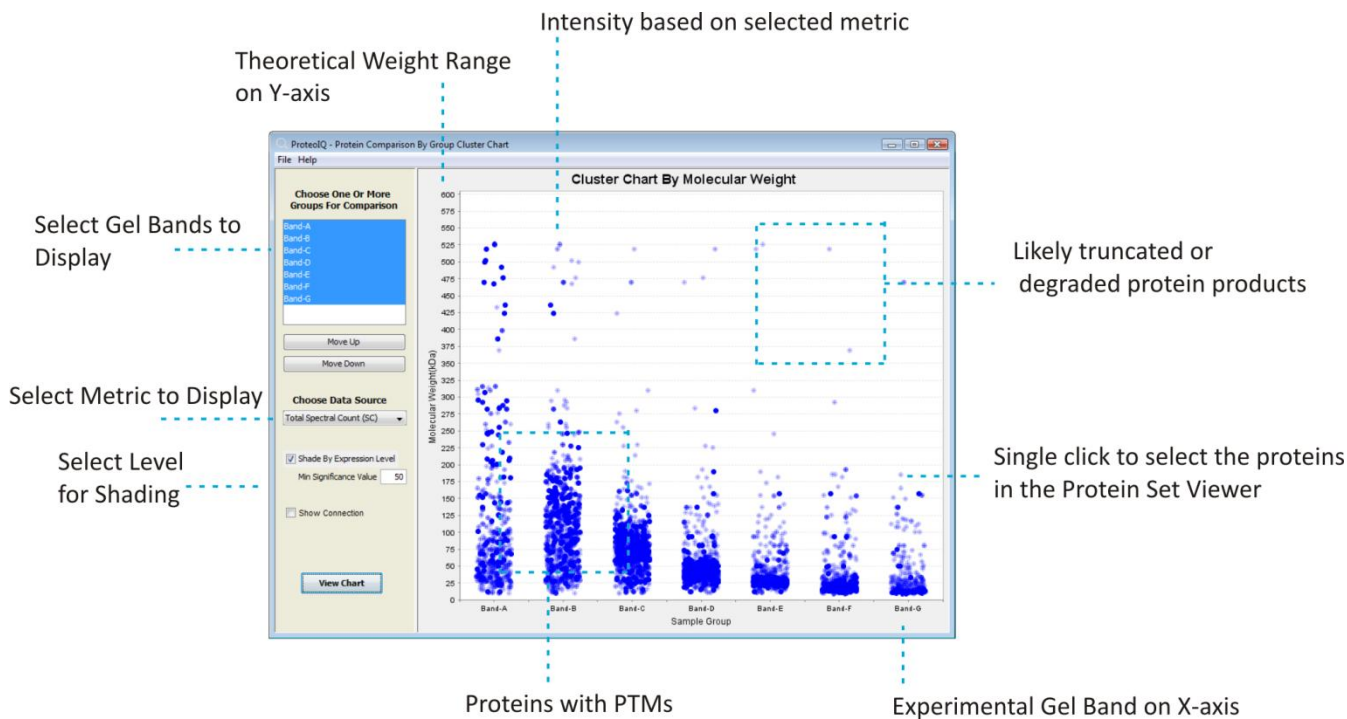
1D Gel View by Protein

The 1D Gel View by Protein (Figure 69) is used to display the results of a 1D GeLC analysis on a protein by protein basis. For a description of how a GeLC experiment should be setup in ProteoIQ, see the [Experimental Approach for GeLC Analysis](#) and the [Preparing Results for GeLC Analysis in ProteoIQ](#) sections.

To Display a 1D Gel View by Protein Plot

1. Select the biological samples (Gel Slices) to display from the Choose One or More Groups for Comparison menu.
2. Choose the metric to display (Note: the darkness of the data point is dependent upon the selected value).
3. Select shading level using the Min Significance Level setting
4. Click View Chart

Figure 69. 1D Gel View by Protein



Interpreting the 1D Gel View by Protein Plot

When gel slices (Biological Samples) are selected they appear on the X-axis. The theoretical molecular weight is displayed on the Y-axis. Each spot is an individual protein and the color intensity is based on the metric chosen in the Choose Data Source drop down menu and the Min Significance Level setting.

In the example shown in Figure 69, Total Spectral Counts is selected and the Min Significance Level was set to 50 spectral counts. The spots colored dark blue have spectral counts exceeding 50 and the lighter colored spots are protein with less than 50 spectral counts.

Double clicking on a data point opens the [Protein Sequence Viewer](#) and single clicking selects the protein in the [Protein Set Viewer](#).

Quantitation

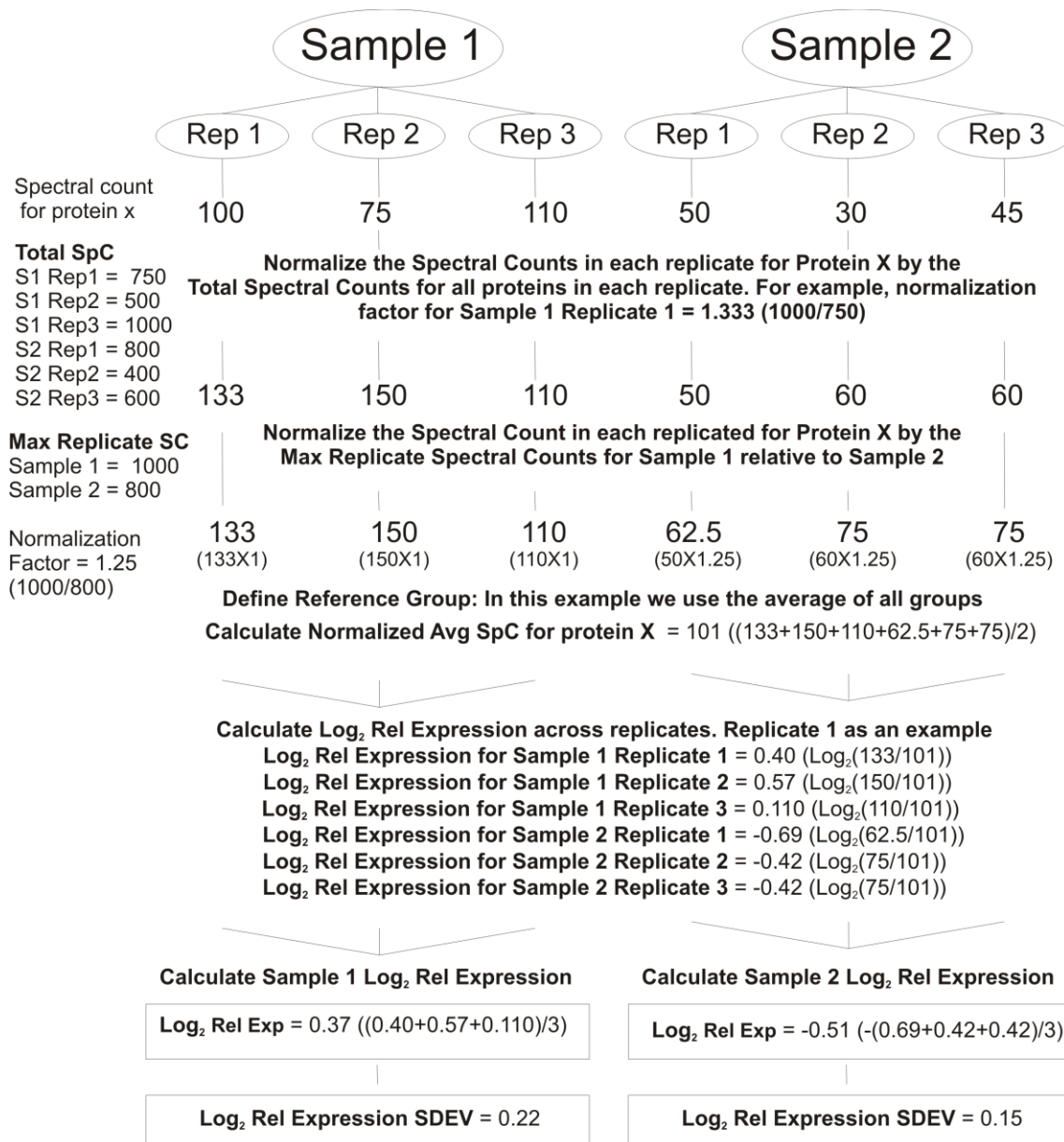
Protein expression can be quantified in ProteoIQ using either [Spectral Counting](#) method or [Reporter Ion Quantitation](#). The following sections describe how protein quantitation is performed in ProteoIQ using the Label Free approach and provides an overview of how to interpret the results.

- 🔍 [Quantitation Overview – Label Free](#)
- 🔍 [Quantitation Without Normalization – Label Free](#)
- 🔍 [Quantitation With Normalization – Label Free](#)
- 🔍 [Modifying Label Free Quantitation Settings](#)
- 🔍 [Quantitation Overview – Reporter Ion Quantitation](#)
- 🔍 [Modifying Reporter Quantitation Settings](#)
- 🔍 [Pseudo Spectral Counts](#)
- 🔍 [Pseudo Spectral Counts Without Normalization](#)
- 🔍 [Pseudo Spectral Counts With Normalization](#)
- 🔍 [Analyzing Replicate Results](#)
- 🔍 [Visualizing Quantitative Results](#)

Quantitation Overview – Label Free

Figure 70 shows a flow chart for how quantitation is performed for a binary comparison of two biological samples with three replicates.

Figure 70. Quantitation Flow Chart – Label Free



The ProteoIQ Quantitation Workflow – Label Free

1. Total Spectral Counts are first determined for each replicate by summing the spectral counts for all proteins identified in that replicate (shown as Total SPC).
2. The spectral counts for each protein in each replicate are then normalized based on Total Spectral Counts in each replicate or using the spectral counts for the selected proteins used in **Control Protein Normalization**.
3. Normalization factors are calculated between Biological Samples. Designated as Normalization Factor using Max Replicate SC.
4. The Replicate Normalized spectral counts are then normalized using the Biological Sample Normalization factor. In the example above, each replicate normalized spectral count is multiplied by the Normalization Factor according to Biological Sample (either 1 or 1.25).
5. Average spectral counts are calculated for every protein across the replicates. Designated as Normalized Avg SpC.
Note: By default expression ratios are calculated relative to the Normalized Avg SpC. The user has the option of changing the reference group using the **Modifying Label Free Quantitation Settings**.
6. Log₂ transformations of the Relative Expression values are calculated for each protein across replicates. Relative Expression values are calculated as the spectral count for protein X divided by the average spectral count for that protein across all biological samples for the same replicate or using the user defined reference group.
7. To determine the Log₂ Relative Expression for a Biological Sample the Replicate Log₂ relative expression values are averaged across all replicates within a biological sample.
8. Standard deviations for the Log₂ Relative Expression are then reported.

Quantitation Without Normalization – Label Free

Relative Expression analysis can be performed using native spectral counts or normalization can be applied. Table 32 contains descriptions of the data types specifically associated with native spectral count quantitation.

Table 32. Native Spectral Count Quantitation

Field	Description
Avg Spectral Count (A-SC)	The average spectral counts across all replicates within a Biological Sample.
A-SC Standard Deviation	The standard deviation of spectral counts between replicates within a Biological Sample.
A-SC % Standard Deviation	The percent standard deviation of spectral counts between replicates within a Biological Sample.
A-SC Log2 Relative Expression	$\text{Log2 (Average SC for protein A in Biological Sample X / Sum of the Average SC for protein A in Biological Sample X, Y, Z...)}$
A-SC Log2 Relative Expression Std Dev	The standard deviation of the Log 2 Expression Values calculated for matched replicates between the Biological Samples.

Quantitation With Normalization – Label Free

Three methods for normalization are available in ProteoIQ. The following sections discuss each method.

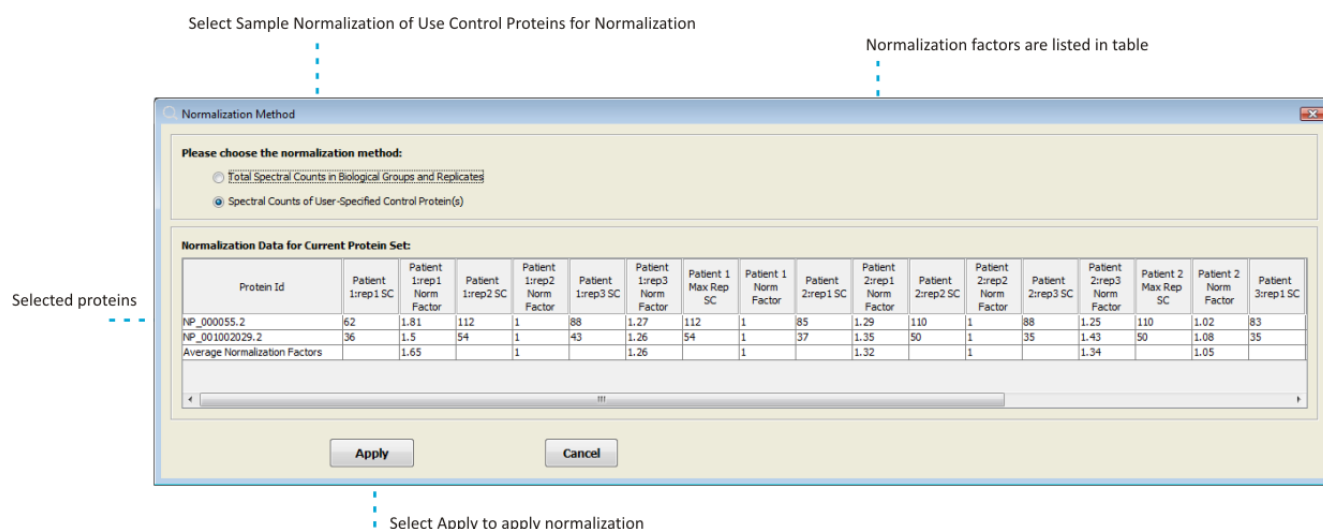
- 🔍 [Normalization by Selected Proteins](#)
- 🔍 [Sample Normalization](#)
- 🔍 [Protein Normalization \(NSAF\)](#)

Normalization by Selected Proteins

Spectral counts for selected proteins (or internal standards) can be used to perform normalization across replicates and biological groups. The following sections describe how to perform normalization by selected proteins (Figure 71).

1. Highlight proteins to use for normalization in the protein set viewer.
2. Right click the selected protein to open the **Pop-Up-Menu**
3. Select Normalize SC From Selected Proteins
4. The Normalization Dialogue Box will appear
5. Select Spectral Counts of User-Specified Control Protein(s) in the Normalization Dialogue Box
6. The normalization factors for each replicate and biological sample will be displayed
7. Click Apply
8. To recalculate Log₂ expression values using the selected normalization factors you **MUST CREATE A NEW PROTEIN SET**
9. To create a new protein set either
 - a. Highlight all of the proteins, right click and select copy to new protein set or
 - b. Select the protein set in the **Filter Previous Set** drop down menu and click Create New Protein Set

Figure 71. Normalization Dialogue Box



To select native or normalized Spectral Count Quantitation select **View > Protein Table Options > Select Columns**.

Table 33 contains descriptions of the data types specifically associated with normalized spectral count quantitation.

Sample Normalization

In many cases over sampling of a proteome of a one biological sample relative to another biological sample can occur. In these cases normalizing the spectral counts for each protein in the under sampled proteome relative to the over sampled state can make the quantitation more accurate. ProteoIQ performs this normalization by comparing the total spectral counts for all proteins identified in each biological group and replicate. The normalization factors are calculated such that the total spectral counts for all proteins in each replicate and biological sample are equal. The normalization factors are then applied to the spectral counts for each protein as shown in Example 2. See [Quantitation Overview](#) for a detailed description of how normalization is applied in ProteoIQ.

Example 2. Normalization Factors

Total Spectral Counts for all proteins within:

Biological Sample 1 = 130,000

Biological Sample 2 = 100,000

Biological Sample 3 = 200,000

Normalization Factors:

Biological Sample 1 = 1.53

Biological Sample 2 = 2.00

Biological Sample 3 = 1.00

Normalization Factors are then applied to individual protein spectral counts.

To view the normalization factors for Sample Normalization, perform the following:

1. Highlight one or more proteins in the protein set viewer.
2. Right click the selected protein to open the [Pop-Up-Menu](#)
3. Select Normalize SC From Selected Proteins
4. The Normalization Dialogue Box will appear
5. Select Total Spectral Counts in Biological Groups and Replicates, in the Normalization Dialogue Box
6. The normalization factors for each replicate and biological sample will be displayed in the Normalization Dialogue box (Figure 71)

To display native or normalized Spectral Count Quantitation select **View > Protein Table Options > Select Columns**.

Table 33 contains descriptions of the data types specifically associated with normalized spectral count quantitation.

Table 33. Normalized Spectral Count Quantitation

Field	Description
Normalized SC (N-SC)	The Average Spectral Counts across all replicates within a Biological Sample that have been multiplied by a Normalization Factor.
N-SC Standard Deviation	The standard deviation of the normalized spectral counts between replicates within a Biological Sample.
N-SC % Standard Deviation	The percent standard deviation of normalized spectral counts between replicates within a Biological Sample.
N-SC Log2 Relative Expression	$\text{Log}_2 \left(\frac{\text{Normalized Average SC for protein A in Biological Sample X}}{\text{Sum of the Normalized Average SC for protein A in Biological Sample X, Y, Z...}} \right)$
N-SC Log2 Relative Expression Std Dev	The standard deviation of the N-SC Log 2 Relative Expression values calculated for matched replicates between the Biological Samples.

Protein Size Normalization (NSAF)

To display Protein Normalized Spectral Counts select: **View > Protein Table Options > Select Columns**

Protein normalized spectral counts (more commonly referred to as normalized spectral abundance factors – NSAF) are calculated as the number of spectral counts for protein x (SpC) divided by the number of amino acids (L) in protein x divided by the sum of SpC/L for all proteins in the experimental dataset.

$$(\text{NSAF})_x = \frac{(\text{SpC}/L)_x}{\sum_{i=1}^N (\text{SpC}/L)_x}$$

L = number of amino acids

SpC = total number of MS/MS spectra identify a protein x

Spectral count normalization based on protein size and protein set composition has been shown to improve the accuracy of spectral count quantification (see references below).

NSAF values have also been applied to study protein complexes and generate protein interaction networks.

References:

Zybailov, BL., Florens, L., Washburn, MP. Quantitative shotgun proteomics using a protease with broad specificity and normalized spectral abundance factors. *Mol. Biosyst.* 2007, 3, 354-360.

Zybailov, BL., Mosley, AL., Sardi, ME., Coleman, MK., Florens, L., Washburn, MP. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome. Res.* 2006, 5, 2339-2347.

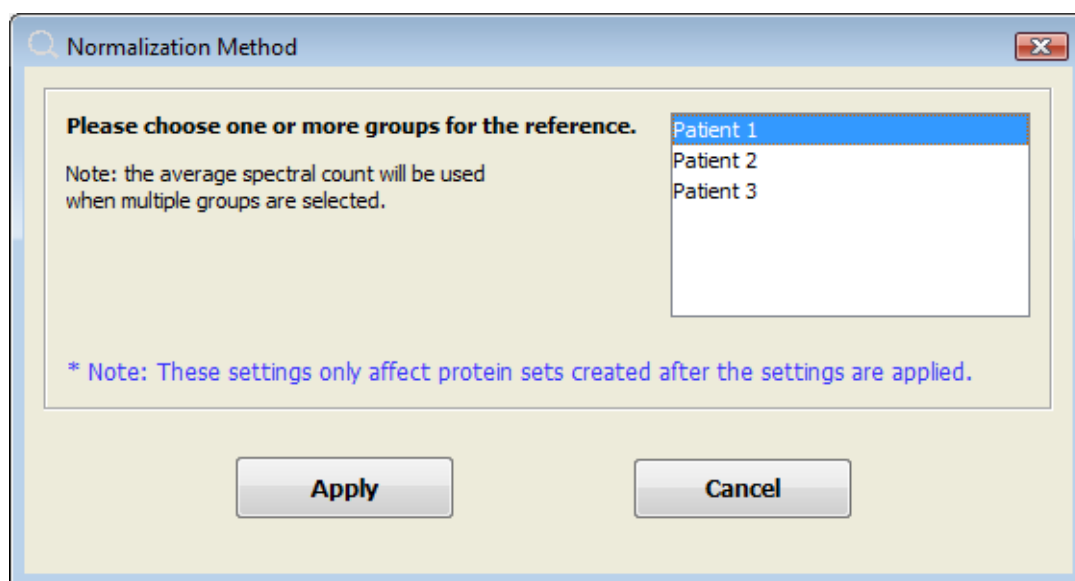
Powell, DW., Weaver, CM., Jennings, JL., McAfee, KJ., He, Y., Weil, PA., Link, AJ. Cluster analysis of mass spectrometry data reveals a novel component of SAGA. *Mol. Cell. Biol.* 2005, 24, 7249-7259.

Modifying Label Free Quantitation Settings

By default, ProteoIQ calculates Log₂ Relative Expression Values by comparing the individual spectral counts from each biological group relative to the average spectral counts across all biological groups. The Modify Label Free Quantitation Settings dialog box allows the user to select the biological group to be used as the reference group for quantitation. For example, in figure 69 there are three biological groups (Patient 1-3). As described in the [Quantitation Overview – Label Free](#) section, the reference group is treated as the average spectral count for protein X across all patients. If patient one is selected as the reference group then the spectral counts for protein X in patient one are treated as the reference and all Log₂ ratios are recalculated and displayed as a new protein set. The following sections describe how to modify label free quantitation settings (Figure 72).

1. Select Edit>Modify Label Free Quant Settings
2. Select the biological group to act as the reference value
3. Click Apply
4. To recalculate Log₂ expression values using the selected biological group as a reference you **MUST CREATE A NEW PROTEIN SET**
5. To create a new protein set either
 - a. Highlight all of the proteins, right click and select copy to new protein set or
 - b. Select the protein set in the [Filter Previous Set](#) drop down menu and click Create New Protein Set

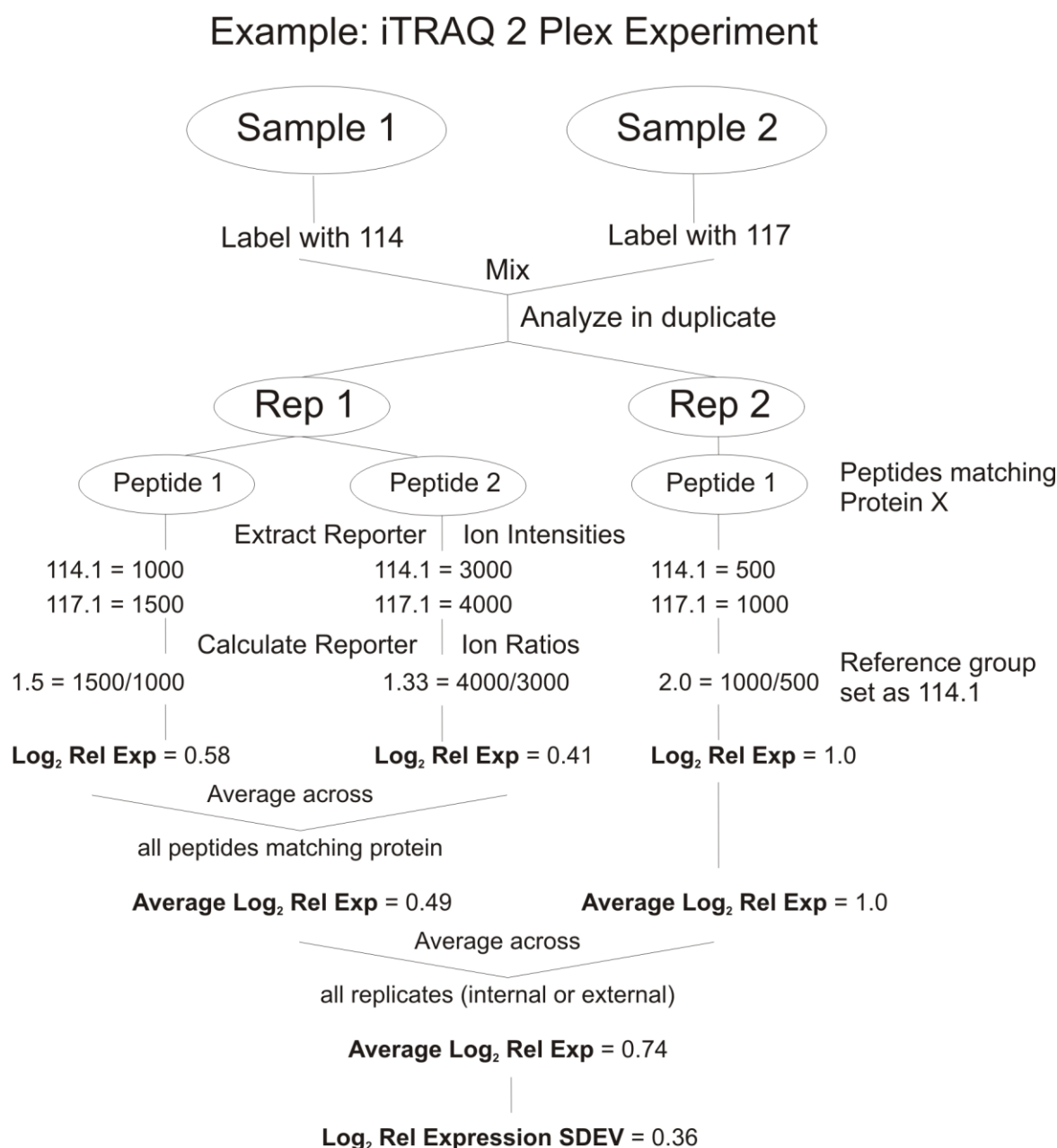
Figure 72. Modify Label Free Quant Settings Dialogue Box



Quantitation Overview – Reporter Ion Quantitation

Figure 73 shows a flow chart for how quantitation is performed for a binary comparison of two biological samples with two replicates using an iTRAQ 2 Plex reagent. Table 34 contains descriptions of the data types specifically associated with Reporter Ion Quantitation

Figure 73. Quantitation Flow Chart – Reporter Ion Quantitation



The ProteoIQ Quantitation Workflow – Reporter Ion Quantitation

1. For every peptide, the reporter ions are detected based upon the user specified Mass Window and Reporter Method. Note: for a reporter ion to be detected, the ion must be within the Mass Window specified and the peptide must contain the modification associated with the reporter label.
2. Reporter ion intensities are then saved for each peptide.
3. Using the Reference Group specified by the user, ratios are calculated by dividing each reporter ion intensity by the intensity of the reference group reporter ion.
4. The ratios are then Log_2 transformed.
5. If multiple peptides are matching a given protein within a single replicate, the Log_2 ratios are average across all peptides within the replicate.
6. Average Log_2 ratios are calculated for every protein across the replicates.
7. Standard Deviations and Percent Standard Deviations are calculated for each protein across all replicates within a Biological Sample.

Table 34. Reporter Ion Quantitation

Field	Description
Ratio	Log2 ratios calculated using reporter ion intensities (specific to protein set viewer)
Log2 Relative Expression	Log2 ratios calculated using reporter ion intensities (specific to the protein sequence viewer)
Log2 Relative Expression Std Dev	The standard deviation of the Log 2 Expression Values calculated for matched replicates between the Biological Samples.

Modifying Reporter Quantitation Settings

During the new project creation, the user must specify settings for selection of reporter ions for isobaric tag based quantitation. These settings include Mass Tolerance and Reference Group. After the project is created the user may modify these settings by selecting the Modify Reporter Quantitation Settings. The following sections describe how to modify reporter quantitation settings (Figure 74).

1. Select Edit>Modify Reporter Quant Settings
2. Set desired Mass Tolerance and Reference Group
3. Click Apply
4. To recalculate Log₂ expression values using the new settings you **MUST CREATE A NEW PROTEIN SET**
5. To create a new protein set either
 - a. Highlight all of the proteins, right click and select copy to new protein set or
 - b. Select the protein set in the [Filter Previous Set](#) drop down menu and click Create New Protein Set

Figure 74. Modify Reporter Quant Settings Dialogue Box

Reporter Quant Settings

Please specify the reporter quantitation settings:

Fragment Mass Tolerance: 0.05

Reference Group: Control

Total Number of Valid Reporter Peptides: 14012 of 14613

* Note: These settings only affect protein sets created after the settings are applied.

Apply Close

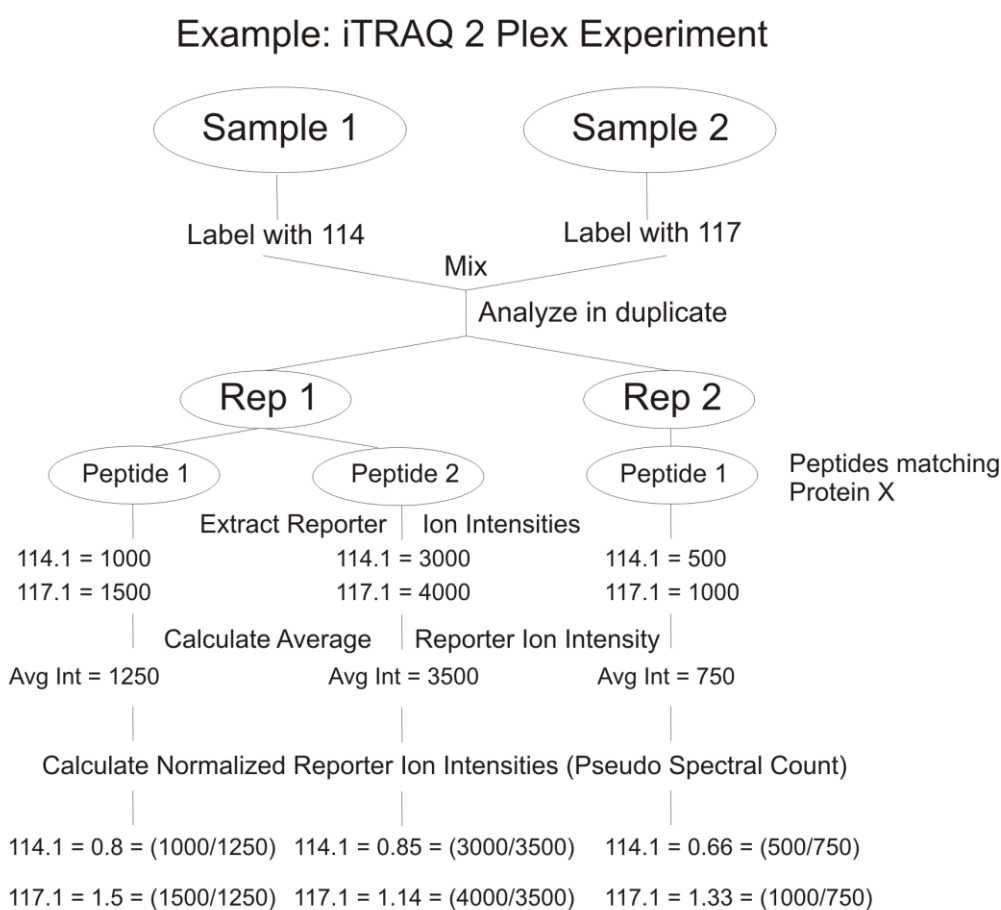
Note: Total Number of Valid Reporter Peptides indicates the number of peptides that have all “detectable” reporter ions within the Fragment Mass Tolerance specified and that have the modification selected in the Select Label Modification page of the New Project Wizard. The peptide must contain all of the reporter ions present. For example, an iTRAQ 2 plex experiment both the reporter ion at 114.1 and 117.1 should be detected to be considered a Valid Reporter Peptide.

Pseudo Spectral Counts

The concept of "pseudo spectral counts" is intended to provide a measure of sampling when a reporter ion based quantitation is performed. Since reporter ion quantitation is a mixture based technique (i.e. all samples are mixed together prior to analysis), a traditional spectral count value is not feasible. To circumvent this, the pseudo spectral count takes into account the number of peptides and relative reporter ion intensities for each biological group to allow a researcher to evaluate a protein's sampling.

Figure 75 shows a flow chart for how pseudo spectral counts are calculated as performed for a binary comparison of two biological samples with two replicates.

Figure 75. Calculation of Pseudo Spectral Counts



The normalized reporter ion intensities represent spectral counts.

For example, for sample one : replicate one, peptide one has a pseudo spectral count = 0.8
At this point the spectral counts are calculated as described under the Label Free Quantitation Section

Note: Once pseudo spectral counts are calculated for an individual peptide, normalization and replicate analysis is performed exactly as described under the Label Free Quantitation section

The Pseudo Spectral Count Workflow

1. For every peptide, the reporter ions are detected based upon the user specified Mass Window and Reporter Method. Note: for a reporter ion to be detected the ion must be within the Mass Window specified and the peptide must contain the modification associated with the reporter label.
2. Reporter ion intensities are then saved for each peptide.
3. The average reporter ion intensity is calculated across all reporter ions for a given peptide.
4. Pseudo spectral counts are then calculated by dividing the intensity of each reporter ion by the average reporter ion intensity.

Pseudo Spectral Counts Without Normalization

Table 35 contains descriptions of the data types specifically associated with native pseudo spectral counts.

Table 35. Native Pseudo Spectral Count Data Types

Field	Description
Pseudo Spectral Count (P-SC)	The average pseudo spectral counts across all replicates within a Biological Sample.
P-SC Standard Deviation	The standard deviation of pseudo spectral counts between replicates within a Biological Sample.
P-SC % Standard Deviation	The percent standard deviation of pseudo spectral counts between replicates within a Biological Sample.

Pseudo Spectral Counts With Normalization

Table 36 contains descriptions of the data types specifically associated with normalized pseudo spectral counts.

Table 36. Normalized Pseudo Spectral Counts

Field	Description
Normalized P-SC (NP-SC)	The Average Pseudo Spectral Counts across all replicates within a Biological Sample that have been multiplied by a Normalization Factor.
NP-SC Standard Deviation	The standard deviation of the normalized pseudo spectral counts between replicates within a Biological Sample.
NP-SC % Standard Deviation	The percent standard deviation of normalized pseudo spectral counts between replicates within a Biological Sample.

Analyzing Replicate Results

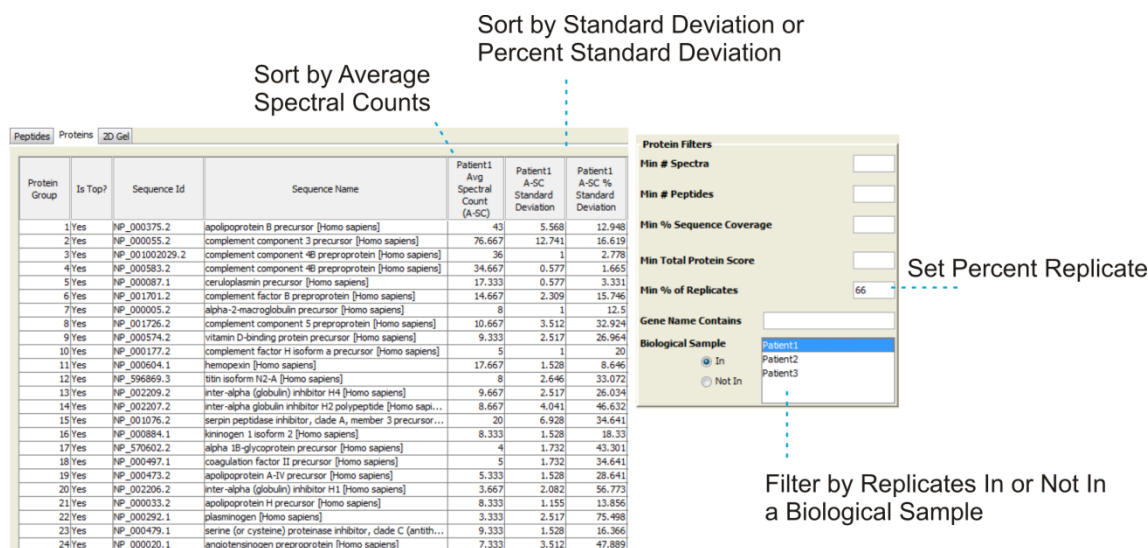
Replicate analyses can be specified in the **Describe Proteome** window of the **New Project Wizard**. When replicates are included, ProteoIQ averages the spectral counts (Avg Spectral Count) across replicates and reports a standard deviation (A-SC Standard Deviation) and percent standard deviation (A-SC % Standard Deviation). The following sections describe how to evaluate replicate results in ProteoIQ:

- 🔍 **Replicate Filters**
- 🔍 **Replicate Plots**

Replicate Filters

Protein identifications can be filtered based on Replicate results using two methods as shown in Figure 76. Using the **Filter Pane**, protein sets can be filtered by **Min Percent Replicates**.

Figure 76. Filtering by Replicates



Setting the **Minimum Percentage of Replicates** is performed as follows:

If three replicates are performed and the user wants to create a protein set containing proteins ONLY identified in 2 of the 3 replicates, then the **Min Percent Replicates** should be set to 66% (2/3). To identify proteins found in at least 1 of 3 replicates set the **Min Percent Replicates** to 33% (1/3).

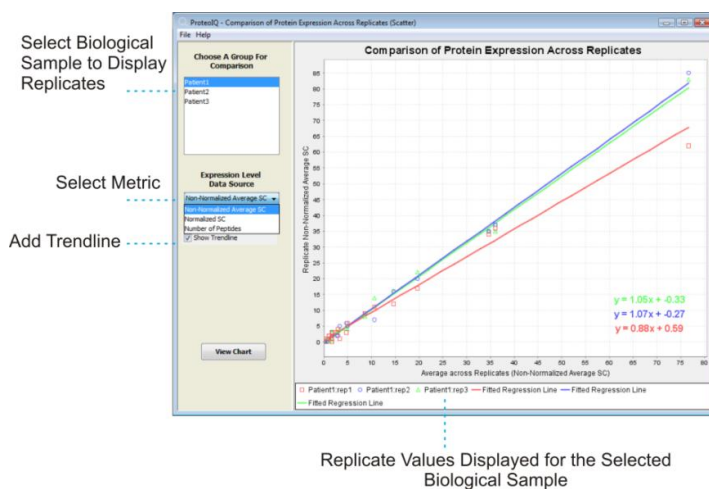
To filter protein sets by Standard Deviation:

1. Click the **Standard Deviation** column in the **Protein Set Viewer** to sort by Standard Deviation
2. Highlight the proteins of interest
3. Right click on the highlighted proteins and select **Copy Selected to New Protein Set**

Replicate Plots

Figure 77 shows the Replicate Analysis (Scatter Plot). To access the Replicate Analysis (Scatter) plot select **Plots > Experimental Reproducibility > Replicate Analysis (Scatter)**.

Figure 77. Replicate Analysis Scatter Plot



To display the plot:

1. Choose a Biological Sample to display in the Choose A Group for Comparison menu
2. Select a data type in the Expression Level Data Source menu
3. Select check box to add a trend line
4. Click View Chart

For the selected Biological Sample, all replicates are displayed in the plot. Each data point represents a protein identified in one or more replicates within the selected Biological Sample. The **y-axis** displays the Spectral Counts (Pseudo Spectral Counts) or Number of Peptides for the protein and the **x-axis** displays the Average Spectral Counts (Average Pseudo Spectral Counts) for the protein calculated across all replicates within the selected Biological Group.

Visualizing Quantitative Results

ProteoIQ includes multiple plots for viewing the results of a quantitative proteome. The following sections describe how to visualize protein expression data in ProteoIQ.

- 🔍 [Quantitative Expression \(Scatter\)](#)
- 🔍 [Quantitative Expression \(Heat Map\)](#)
- 🔍 [Relative Expression Changes \(Cluster Plot\)](#)
- 🔍 [Relative Expression Differences \(Dot\)](#)

Quantitative Expression (Scatter)

The quantitative expression scatter plot is used to compare Spectral Counts (Pseudo Spectral Counts), Peptides, or % Sequence Coverage between two or more Biological Samples (Figure 78). To open the quantitative expression scatter plot: **Plots > Comparative Proteomics > Quantitative Expression (Scatter)**.

Figure 78. Quantitative Expression Scatter Plot

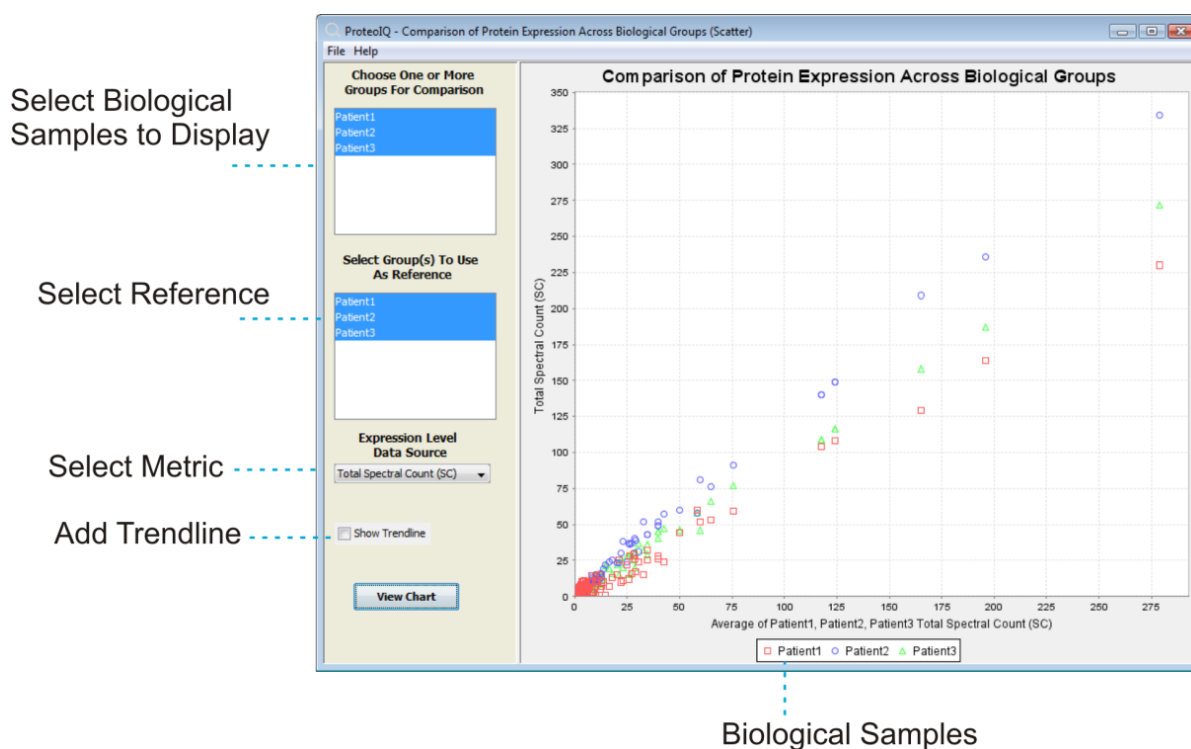


Table 37. Parameters in the quantitative expression scatter plot

Field	Description
Choose One or More Groups for Comparison	Selects Biological Samples to display. When a Biological Sample is selected, the proteins (Data Source, Spectral Counts, # Peptides etc.) will be displayed as individual data points with the selected Data Source on the Y-axis.
Select Groups to use as Reference	Selects Biological Samples to display proteins (Data Source, Spectral Counts, # Peptides etc.) as a reference on the X-axis. When more than one Biological Sample is selected as a reference the Data Source is averaged for each protein between the selected Biological Samples.
Expression Level Data Source	The Data Source specifies when type of data to plot for the proteins. Options are Total Spectral Counts, Non-Normalized Average Spectral Counts, Normalized Spectral Counts, Number of Peptides, Percent Sequence Coverage. Note: for reporter quantitation the options are Non-Normalized Pseudo Spectral Counts and Normalized Pseudo Spectral Counts
Show Trend line	Calculates a linear trend line through each data series for all Biological Samples.

Quantitative Expression (Heat Map)

The quantitative expression heat map is used to compare Spectral Counts, Peptides, or % Sequence Coverage between two or more Biological Samples (Figure 79). To open the quantitative expression heat map: **Plots > Comparative Proteomics > Quantitative Expression (Heat Map)**. The Heat Map is similar to standard Heat Maps for microarray analysis in which the Log2 expression values are color coded based on relative expression. The colors range from red (lowest expression) to green (highest expression). Absolute values (spectral counts, peptides etc.) can also be selected and the number of color partitions can be set.

Figure 79. Quantitative Expression Heat Map

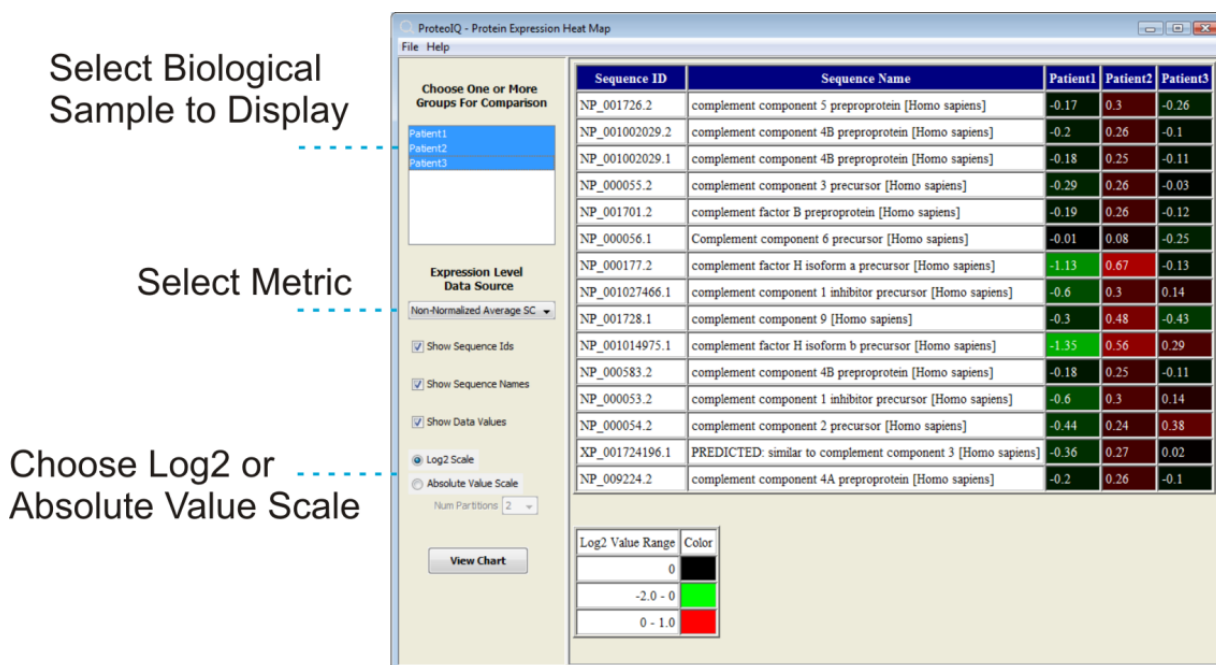


Table 38. Quantitative Expression Heat Map

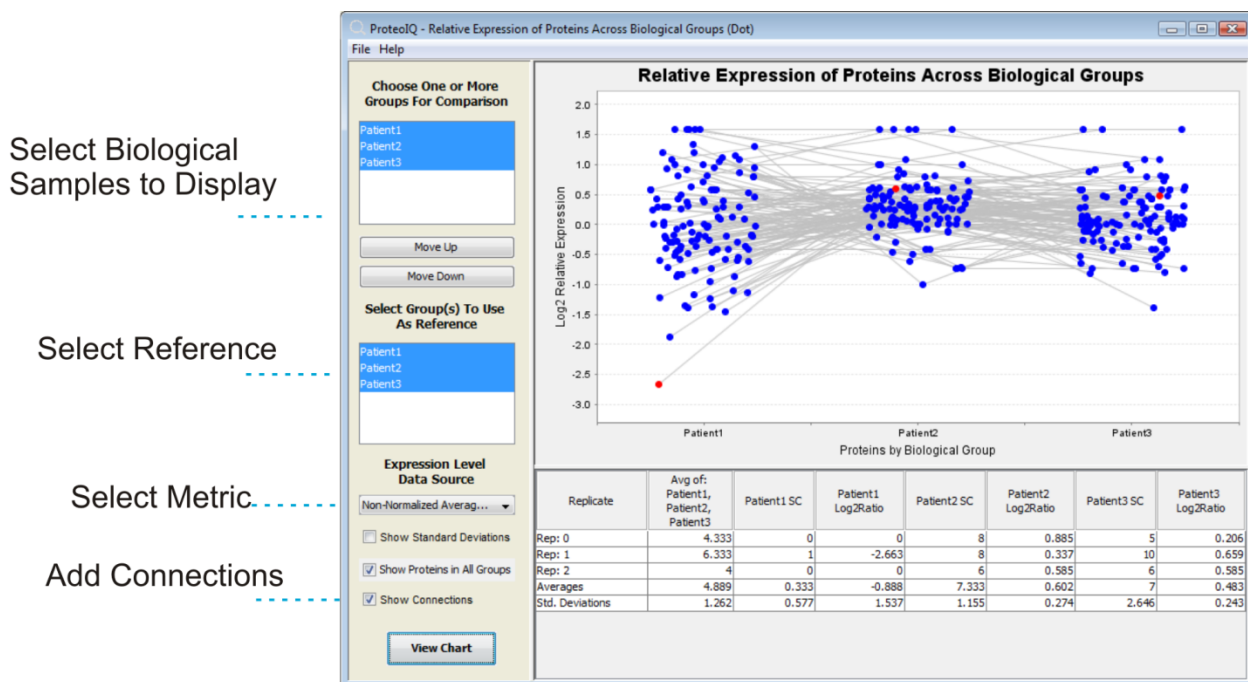
Field	Description
Choose One or More Groups for Comparison	Selects Biological Samples to display as columns in the heat map. When a Biological Sample is selected, the proteins (Data Source, Spectral Counts, # Peptides etc.) will be displayed in tabular format underneath the Biological Sample column header.
Expression Level Data Source	The Data Source specifies when type of data to display for the proteins. Options are Total Spectral Counts, Non-Normalized Average Spectral Counts, Normalized Spectral Counts, Number of Peptides, Percent Sequence Coverage. Note: for reporter quantitation the options are Non-Normalized Pseudo Spectral Counts and Normalized Pseudo Spectral Counts
Log2 Scale	Displays Log2 expression values calculated using the selected Data Source.
Absolute Scale	Displays the native Total Spectral Counts, Non-Normalized Average Spectral Counts, Normalized Spectral Counts, Number of Peptides, Percent Sequence Coverage. Note: for reporter quantitation the options are Non-Normalized Pseudo Spectral Counts and Normalized Pseudo Spectral Counts
Number of Partitions	Sets the number of color gradients. For example if the Number of Partitions is set to 4 and the maximum spectral counts are 100 and the minimum spectral counts is 0. Then the color partitions will be set as follows: $(100-0)/4 = 25$. The first color will range from 1-25, the second from 26-50 etc.

Relative Expression Changes (Cluster Dot)

The relative expression changes (cluster dot) plot is used to compare Log₂ Relative Expression Values between two or more Biological Samples (Figure 80). To open the Relative Expression Changes (Cluster Dot) Plot: **Plots > Comparative Proteomics > Relative Expression Changes (Cluster Dot)**. The Cluster Dot Plot provides an overview of expression for all proteins from the selected Biological Samples.

Biological Samples are listed on the X-axis and the Y-axis displays Log₂ Relative Expression.

Figure 80. Relative Expression Changes (Cluster Dot) Plot



Replicate Data Displayed for Selected Protein

Table 39. Parameters in the Relative Expression Changes (Cluster Dot) Plot

Field	Description
Choose One or More Groups for Comparison	Selects Biological Samples to display. When a Biological Sample is selected, the Log ₂ Expression for proteins (using either Non-Normalized Average Spectral Counts, or Normalized Spectral Counts.) will be displayed as data points with the selected Data Source on the Y-axis.
Select Groups to use as Reference	Determines the Biological Samples that act as the reference for the Log ₂ Relative Expression calculations. For example, if Biological Sample 1 is selected as Comparison Group and Biological Sample 2 is selected as Reference Group the Log ₂ value will be calculated as follows: Log ₂ (SpC in BioGroup1 /SpC in BioGroup 2). When more than one Biological Sample is selected as a reference the Data Source is averaged for each protein between the selected Biological Samples.
Expression Level Data Source	The Data Source specifies when type of data to use for the Log₂ relative expression calculation. Options are Non-Normalized Average Spectral Counts, or Normalized Spectral Counts. Note: for reporter quantitation the Data Source field is deactivated and expression levels are calculated as described Quantitation Overview – Reporter Ion Quantitation
Show Standard Deviations	Adds Log ₂ Standard Deviations to the plot.
Show Proteins in All Groups	Hides proteins that are NOT present in all the selected Biological Samples.
Show Connections	Adds lines between common proteins found in more than one group.

Relative Expression Differences (Dot)

The relative expression differences (cluster dot) plot is used to compare DIFFERENCE between the $\log_2$ Relative Expression Values between two or more Biological Samples (Figure 81). While the Relative Expression Changes (Cluster Dot) Plot displays the $\log_2$ expression values for proteins in multiple Biological Samples, the Relative Expression Differences Plot asked the question

“What is the difference in $\log_2$ Relative Expression between two or more Biological Samples?”

To open the Relative Expression Differences (Dot) Plot: **Plots > Comparative Proteomics > Relative Expression Differences (Dot).**

Figure 81. Relative Expression Differences (Dot) Plot

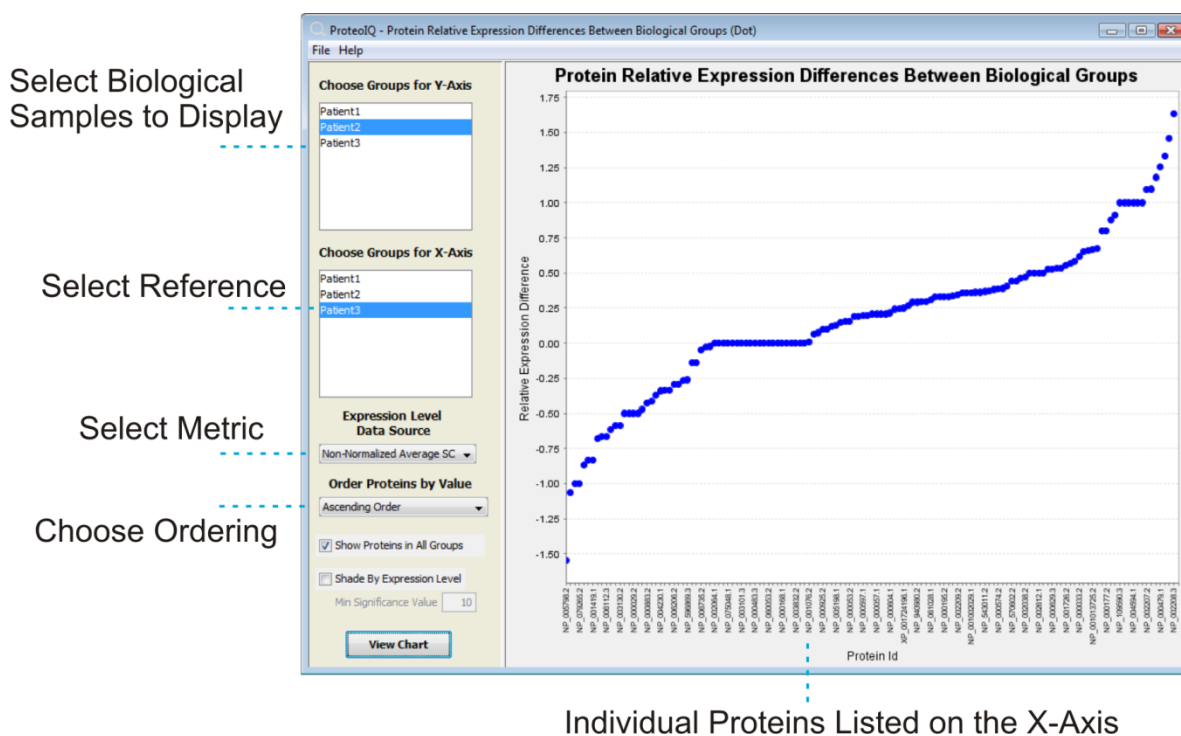


Table 40. Parameters in the Relative Expression Differences (Dot) Plot

Field	Description
Choose Groups for Y-Axis	When a Biological Sample is selected, the Log ₂ Expression Difference for that samples proteins (using either Non-Normalized Average Spectral Counts, Normalized Spectral Counts, or Reporter Ion Quantitation.) will be displayed as data points.
Choose Groups for X-Axis	Determines the Biological Samples that act as the reference for the Log ₂ Relative Expression Difference calculations. For example, if Biological Sample 1, proteinX (Log ₂ = 0.5) is selected as Y-Axis and Biological Sample 2, proteinX (Log ₂ = -0.5) is selected as Reference Group the Log ₂ Expression Difference will be calculated as follows: $1.0 = 0.5 - (-0.5)$. When more than one Biological Sample is selected as the reference (X-axis) the Data Source is averaged for each protein between the selected Biological Samples prior to the Log ₂ calculation being performed.
Expression Level Data Source	The Data Source specifies when type of data to use for the Log₂ relative expression calculation. Options are Non-Normalized Average Spectral Counts, or Normalized Spectral Counts. Note: for reporter quantitation the Data Source field is deactivated and expression levels are calculated as described Quantitation Overview – Reporter Ion Quantitation
Order Protein by Value	Select either Descending or Ascending to modify the order of the Log ₂ Relative Expression Difference values.
Show Proteins in All Groups	Hides proteins that are NOT present in all the selected Biological Samples.
Shade by Expression Value	Sets the color intensity for the data points based on the metric chosen in the Choose Data Source drop down menu and the Min Significance Level setting. For example if Normalized Spectral Counts is selected and the Min Significance Level was set to 50. The spots colored dark blue would have normalized spectral counts exceeding 50 and the lighter colored spots are protein with less than 50 spectral counts.

Other Topics

- 🔍 [Protein Grouping](#)
- 🔍 [Source](#)
- 🔍 [Redundancy and Multiple Charge States](#)
- 🔍 [Discriminant Score \(F value\)](#)

Protein Grouping

Correlating peptides to proteins in a large-scale proteomics project is not a straightforward process, because protein databases contain many homologous proteins which are often indistinguishable by single peptide identification. To account for this, it has become readily accepted that peptide identifications should be distributed into protein groups. A protein group represents a set of peptides used to identify a group of homologous proteins.

In the example below five peptides were identified:

APHLRK – peptide A

EIEQLFEHHHQK – peptide B

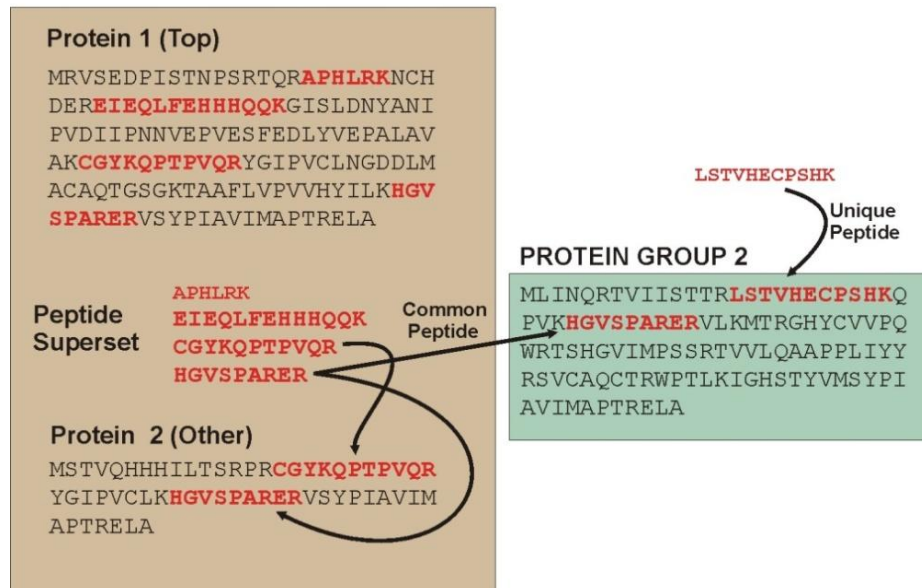
CGYKQTPVQR – peptide C

HGVSPARER – peptide D

LSTVHECPSHK – peptide E

Peptides (A,B,C,D) all map to protein 1. However, peptides (C, D) are also observed on protein 2; therefore, proteins 1 and 2 are grouped as PROTEIN GROUP 1, with protein 1 being listed as TOP, since it is able to explain the presence of all four peptides - whereas protein 2 can only account for two. The identification of peptide E allows a unique PROTEIN GROUP 2 to be formed, because peptide E is not observed in the sequence of protein 1 or 2.

PROTEIN GROUP 1



For a more detailed discussion on the formation of protein groups and their importance in the calculation of false discovery rates or probabilities see the following references:

References:

Weatherly, DB, Atwood, JA 3rd, Minning, TA, Cavola, C, Tarleton, RL, Orlando, R. A Heuristic method for assigning a false-discovery rate for protein identifications from Mascot database search results.

Molecular and Cellular Proteomics. 2005, 4, 762-72.

Nesvizhskii, AI, Keller, A, Kolker, E, Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 2003, 75, 4646-4658.

Nesvizhskii, AI, Aebersold, R. Analysis, statistical validation and dissemination of large-scale proteomics data sets generated by tandem MS. *Drug. Discov. Today*. 2004, 4, 173-181.

Zhang, B, Chambers, MC, Tabb, DL. Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. *J Proteome Res*. 2007, 6, 3549-57.

Source

TOP indicates that the protein sequence can account for all peptides within a PROTEIN GROUP.

OTHER indicates that the protein is a member of a PROTEIN GROUP but it is not the TOP. Protein assignments list as OTHER contain a subset of peptides that were observed in the TOP identification but cannot be distinguished as unique proteins because of shared peptide representation.

Note: Where a protein is listed as OTHER does not mean that the protein was not identified. It simply means that the peptides observed for that protein do not allow it to be distinguished from the other proteins within the group. It is always plausible that several or all of the members of a protein group are expressed.

Redundancy and Multiple Charge States

While spectra from multiple precursor ion charge states may exhibit distinct fragmentation patterns, these peptides are not considered as independent peptide matches by ProteoIQ and are not allowed to contribute to the total score. We have adopted this approach because of the fact that during database searching precursor ions of multiple charge states will be converted to their singly charged precursor masses prior to comparison with the *in silico* derived peptides. The subset of *in silico* derived peptides passing within the precursor ion mass tolerance will essentially be identical for the singly charged and multiply charge species. Thus identical assignments made from peptides with different precursor ion charge states are not unequivocally independent events. Conversely, peptide assignments with alternate modifications are matched through different parent mass filters and are treated as independent.

Discriminant Score (F value)

F value calculation for Mascot®

$$F \text{ value} = co + c1(\text{IonScore} - \text{IdentityScore} + \text{AvgIdentityScore})$$

Example:

Peptide - IAYNVEAAR

Ion score = 58.4

Identity score = 25

Average Identity score = 20

$$F \text{ value} = -3.0 + 0.1(58.4 - 25 + 20)$$

$$F \text{ value} = 2.34$$

F value calculation for SEQUEST®

$$Xcorr' = \frac{\ln(Xcorr)}{\ln(\#AA)}$$

Xcorr' – Corrects for peptide length dependence in Xcorr

$$F \text{ value} = 8.362(Xcorr') + 7.386(\Delta Cn) - 0.194(\ln[\text{SpRank}]) - 0.314(\Delta M) - 0.959$$

Example:

Peptide - LLLEFDISK

Xcorr = 2.823

$\Delta Cn = 0.1$

SpRank = 692.7

$\Delta M = 0.4$

$$F \text{ value} = 8.362(0.472) + 7.386(0.1) - 0.194(6.54) - 0.314(0.4) - 0.959$$

$$F \text{ value} = 2.33$$

As described in Keller *et al*, ProteoIQ calculates a discriminant score (F value) for every peptide assignment to a given spectra. This allows the assignment of a single score for peptide assignments from different database search engines and thus different scoring systems. For Mascot search results F values are calculated using the Ion Score, Identity Score and Average Identity Score. For SEQUEST results, the Xcorr are first normalized (Xcorr') to account for the Xcorr dependence on peptide size. F values are then calculated using the Xcorr', ΔCn , SpRank, and ΔM . F values can be employed in ProteoIQ to calculate peptide probabilities and/or false discovery rates.

References:

Keller, A., Nesvizhskii, A. I., Kolker, E., Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 2002, 74, 5383-5392.

Legal Notices

Copyright

© 2009 BIOINQUIRE, LLC. All rights reserved.

This manual and the software described in it are the exclusive property of BIOINQUIRE, LLC. BIOINQUIRE, LLC (including BIOINQUIRE, LLC officers, employees, agents, directors, independent contractors, affiliates, distributors and successors) has not made or granted any express warranties concerning the information contained in this manual. The content of this manual are for information purposes only and is subject to change at any time. BIOINQUIRE, LLC shall not be liable for any direct, indirect, lost profits, consequential, exemplary, incidental or punitive damages, regardless of the form of action, as a result of any errors or inaccuracies in this manual.

No part of this manual, including text and images shall be copied without the prior written consent of BIOINQUIRE, LLC.

Microsoft, Windows and Vista are registered trademarks of Microsoft Corporation. Mascot is a registered trademark of Matrix Science Limited Corporation. SEQUEST is a registered trademark of the University of Washington. Java is a registered trademark of Sun Microsystems, Inc. iTRAQ is a registered trademark of Applied Biosystems. Tandem Mass Tags is a registered trademark of Proteome Sciences.

Index

- # peptide, 70
- # Spectra, 70
- % of Replicates, 70
- % Sequence Coverage, 70
- .dat, 37, 42
- .out, 42
- accession number, 80
- accession numbers, 62
- All parsed peptides from results files, 86
- Available Results Files viewer, 46, 57, 58
- biological sample, 41, 42, 43, 70, 117
- Biological Sample, 70
- Charge state, 69
- Charge State, 21, 76, 77, 90, 91
- Clear Selected, 24
- Cluster parsed peptides to my protein database, 45
- Combine all results files (MudPIT), 42
- Comparison Function, 16, 73
- Copy, 24
- Copy to New Protein Set, 24
- Create ion data for ALL peptides, 48, 61, 63, 69, 70
- Create ion data for top scoring, 48, 61, 63, 68, 69, 70, 71, 72, 74, 76, 77, 78, 80, 81, 82, 86, 87, 88, 96, 97, 105, 106
- Create no ion data, 48, 61, 63, 68, 69, 70, 71, 72, 74, 76, 77, 78, 79, 80, 81, 82, 86, 87, 88, 96, 97, 105, 115, 118, 125, 126, 130, 132, 134, 136
- Database name, 61
- Database type, 61
- decoy database, 42, 44, 61, 71, 72
- Describe Database, 44
- Describe Proteome, 43
- Describe Samples, 43, 56
- Description, 61, 63
- Difference, 74
- discriminant score, 141
- Discriminant score, 72
- Discriminate score, 71
- Each result file represents an individual sample, 42
- Enzyme specificity, 69
- Export, 23, 30, 85, 86
- Expression Ratios, 17
- Extract ion data for peptides id'd from a single spectrum, 48, 61, 63, 68, 69, 70, 71, 72, 74, 76, 77, 78, 79, 80, 81, 82, 86, 87, 88, 96, 97, 105, 106, 107, 115, 118, 125, 126, 130, 134, 136
- F value, 141
- false discovery rate, 29, 67, 71, 72
- False Discovery Rate, 29, 48, 89, 94, 100, 103, 104
- False Discovery Rates, 42
- FASTA Database, 37, 38
- Fragment Ion Type, 21
- Fragment mass tolerance, 87, 96
- Fval, 71
- Gene name contains, 70
- Hyperlink, 65
- Installing ProteoIQ, 8
- intensity percentile, 87, 88, 96, 97, 105, 106
- Intersection, 74, 76
- Ion Series, 87, 96
- Label peaks with mass to charge, 88, 97
- Label regardless of intensity, 88, 97
- Licensing, 8
- Mascot, 10, 30, 37, 42, 47, 49, 69, 71, 77, 78, 86, 90, 141, 142
- Max Peptide FDR, 71
- Max Protein FDR, 71
- Metric to use for FDR, 71
- Min peptide length, 50
- Min score, 48, 49
- Min score differential, 49
- Modification, 20, 70
- Modification Name, 70
- Modify Database Wizard, 10, 11, 14, 18, 36, 38, 60, 61
- Modify Databases, 44
- Neutral Losses, 21, 88, 95, 97, 98, 99
- New Project Wizard, 12, 13
- non-target database, 42
- NSAF, 119
- OTHER, 139
- Parse Files Progress, 50
- Peptide Length, 69
- Peptide probability, 71
- Peptide Probability, 26, 29, 48, 72, 77, 89, 91
- Peptide Table Options, 20, 25, 75, 76

- Peptide/Protein Viewer, 15, 39, 52
- PeptideProphet, 29
- Probability, 29
- project name, 40
- Properties, 23, 84
- Protein groups per peptide, 69
- Protein ID parse rule, 63
- Protein list html, 86
- Protein Name parse rule, 63
- protein names, 62
- Protein Normalized, 119
- Protein Probability, 29, 48, 72, 80, 100, 101
- Protein Sequence Viewer, 15, 17, 18, 19, 21, 28, 75, 83, 95, 112
- Protein sequences, 86
- Protein sets to export, 85
- ProteinProphet, 29
- Proteins per peptide, 69
- ProValT, 71
- regular expression, 63
- relative standard deviation, 81, 82
- Remove, 74
- Rename, 74
- Replicates, 43
- Sample Names, 43, 56
- Save, 14, 23, 33, 34, 60, 62
- Score, 17, 47, 49, 67, 69, 71, 76, 77, 80, 89, 90, 91, 92, 93, 94, 100, 102, 137, 141
- Select All, 24, 25
- Select by Find, 24, 25
- Select by List, 24
- Select database, 61
- Select Existing, 63
- Select Files for Parsing, 59
- SEQUEST, 10, 37, 38, 42, 45, 47, 49, 69, 71, 77, 78, 86, 90, 141, 142
- Source, 139
- Specify Parse Rules, 63
- Spectra per peptide, 69
- spectral count, 48, 69, 119
- Spectral Count, 17, 77, 80, 81, 82, 105, 115, 116, 117, 125, 127
- target, 29, 34, 42, 44, 61, 71, 72
- TOP, 80, 139
- Total protein score, 70
- total score, 70
- Total Selected Files, 46
- Tutorials, 6
- Union, 74
- Validate Proteome, 23, 33, 34
- Validated Project, 34
- View Chart, 96
- Whole Project, 34
- X!Tandem, 10, 30, 37, 38, 47, 49, 69, 71, 77, 78, 90

BIOINQUIRE

220 Riverbend Rd
Suite 103
Athens, GA 30602

Phone 706.583.4000

Fax 706.583.4000

www.bioinquire.com